Business Management
Study Manuals

Diploma in
Business Management

ECONOMIC PRINCIPLES AND THEIR APPLICATION TO BUSINESS

The Association of Business Executives

5th Floor, CI Tower • St Georges Square • High Street • New Malden
Surrey KT3 4TE • United Kingdom
Tel: + 44(0)20 8329 2930 • Fax: + 44(0)20 8329 2945
E-mail: info@abeuk.com • www.abeuk.com

© **Copyright, 2008**

The Association of Business Executives (ABE) and RRC Business Training

All rights reserved

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, electrostatic, mechanical, photocopied or otherwise, without the express permission in writing from The Association of Business Executives.

Diploma in Business Management

ECONOMIC PRINCIPLES AND THEIR APPLICATION TO BUSINESS

Contents

Unit	*Title*	*Page*
1	**The Economic Problem and Production**	**1**
	Introduction to Economics	2
	Basic Economic Problems and Systems	4
	Nature of Production	6
	Production Possibilities	11
	Some Assumptions Relating to the Market Economy	14
2	**Consumption and Demand**	**17**
	Utility	18
	The Demand Curve	21
	Utility, Price and Consumer Surplus	24
	Individual and Market Demand Curves	25
3	**Demand and Revenue**	**27**
	Influences on Demand	29
	Price Elasticity of Demand	33
	Further Demand Elasticities	36
	The Classification of Goods and Services	38
	Revenue and Revenue Changes	41
4	**Costs of Production**	**49**
	Inputs and Outputs: Total, Average and Marginal Product	50
	Factor and Input Costs	56
	Economic Costs	65
	Costs and the Growth of Organisations	66
	Small Firms in the Modern Economy	69
5	**Costs, Profit and Supply**	**75**
	The Nature of Profit	76
	Maximisation of Profit	79
	Influences on Supply	86
	Price Elasticity of Supply	92

Unit	*Title*	*Page*
6	**Markets and Prices**	**101**
	Nature of Markets	103
	Functions of Markets	105
	Prices in Unregulated Markets	106
	Price Regulation	110
	Defects in Market Allocation	112
	The Case for a Public Sector	116
	Methods of Market Intervention: Indirect Taxes, Subsidies and Market Equilibrium	117
	Using Indirect Taxes and Subsidies to Correct Market Defects	122
7	**Market Structures: Perfect Competition versus Monopoly**	**129**
	Meaning and Importance of Competition	130
	Perfect Competition	131
	Monopoly	137
8	**Market Structures and Competition: Monopolistic Competition and Oligopoly**	**145**
	Monopolistic Competition	146
	Oligopoly	148
	Profit, Competition, Monopoly, Oligopoly and Alternative Objectives for the Firm	154
9	**The National Economy**	**159**
	National Product and its Measurement	160
	National Product	166
	National Expenditure	169
	National Income	170
	Equality of Measures	172
	Use and Limitations of National Income Data	173
	National Product and Living Standards	176
10	**Determination of National Product: The Keynesian Model of Income Determination and the Multiplier**	**179**
	Changes in Consumption, Saving and Investment	180
	Government Spending and Taxation	184
	Changes in Equilibrium, the Multiplier and Investment Accelerator	185
	The Role of the Government in Income Determination: the Government's Budget Position and Fiscal Policy	192
11	**Macroeconomic Equilibrium and the Deflationary and Inflationary Gaps**	**195**
	National Income Equilibrium and Full Employment	196
	The Basic Keynesian View	196
	The Deflationary Gap	197
	The Inflationary Gap	200
	The Aggregate Demand/Aggregate Supply Model of Income Determination	203
	Financing Fiscal Policy: Budget Deficits and Public Sector Borrowing	211
	The Limitations of Fiscal Policy	214

Unit	*Title*	*Page*
12	**Money and the Financial System**	**217**
	Money in the Modern Economy	218
	The Financial System	220
	The Banking System and the Supply of Money	224
	The Central Bank	226
	Interest Rates	228
13	**Monetary Policy**	**233**
	Options for Holding Wealth	234
	Liquidity Preference and the Demand for Money	236
	Implications of the Interest Sensitivity of the Demand for Money	238
	Changes in Liquidity Preference	241
	The Quantity Theory of Money and the Importance of Money Supply	242
	Methods of Controlling the Supply of Money	244
	Monetary Policy and the Control of Inflation	245
14	**Macroeconomic Policy**	**249**
	The Major Economic Problems	250
	Policy Instruments Available to Governments	253
	Policy Conflicts and Priorities	258
	Supply-side Policies	259
15	**The Economics of International Trade**	**265**
	Gains from Trade and Comparative Cost Advantage	266
	Trade and Multinational Enterprise	269
	Free Trade and Protection	272
	Methods of Protection	276
	International Agreements	279
16	**National Product and International Trade**	**285**
	International Trade and the Balance of Payments	286
	Balance of Payments Problems, Surpluses and Deficits	293
	Balance of Payments Policy	297
17	**Foreign Exchange**	**301**
	International Money	302
	Exchange Rates and Exchange Rate Systems	304
	Exchange Rate Policy	310
	Macroeconomic Policy in Open Economy	311

Study Unit 1

The Economic Problem and Production

Contents ***Page***

Introduction to Economics **2**

A. Basic Economic Problems and Systems **4**
Some Fundamental Questions 4
Choice and Opportunity Cost 5

B. Nature of Production **6**
Economic Goods and Free Goods 6
Production Factors 6
Enterprise as a Production Factor 7
Fixed and Variable Factors of Production 8
Production Function 8
Total Product 9

C. Production Possibilities **11**

D. Some Assumptions Relating to the Market Economy **14**
Consistency and Rationality 14
The Forces of Supply and Demand 14
Basic Objectives of Producers and Consumers 15
Consumer Sovereignty 15

© ABE and RRC

How to Use the Study Manual

Each study unit begins by detailing the relevant syllabus *aim* and *learning outcomes or objectives* that provide the rationale for the content of the unit. For this unit, see the section below. You should commence your study by reading these. After you have completed reading each unit you should check your understanding of its content by returning to the objectives and asking yourself the following question: *"Have I achieved each of these objectives?"*

To assist you in answering this question each unit in this subject ends with a list of *review points*. These relate to the content of the unit and if you have achieved the objectives or learning outcomes you should have no trouble completing them. If you struggle with one or more, or have doubts as to whether you really do understand some of the key concepts covered, you should go back and reread the relevant sections of the unit. Ideally, you should not proceed to the next unit until you have achieved the learning objectives for the previous unit. Your tutor should be able to assist you in confirming that you have achieved all the required objectives.

Objectives

The aim of this unit is to explain the problem of scarcity, the concept of opportunity cost, the difference between macroeconomics and microeconomics and the difference between normative and positive economics.

When you have completed this study unit you will be able to:

- explain the problems of scarcity and opportunity cost
- explain how scarcity and opportunity cost are related using numerical examples and a production possibility frontier
- explain what is meant by free market, command and mixed economies
- discuss, using real world examples, the relative merits of these alternative regimes
- explain what is meant by microeconomics and macroeconomics and discuss the differences between these areas
- explain the meaning and implications of the *ceteris paribus* assumption in microeconomics
- explain what is meant by normative and positive economics and discuss the differences between these terms.

INTRODUCTION TO ECONOMICS

The study of economics is important because we all live in an economy. Our well-being is closely related to the success, or otherwise, of both the economy in which we live and that of all the other economies in the world. Whether people have jobs or are unemployed, the kind of work people do, the things they produce, how much they are paid, what they purchase, how much they consume, and the influence of the government on economic activity are the subject matter of economics. The study of economics is important for a proper understanding of business. This is because we are all consumers and will be workers for a large part of our lives, so that what we do determines how well business does. The study is important for business because often common sense is not a good guide to how a firm should operate to get the best out of a particular situation. What the study of economics reveals is that in many situations what is obvious is not always correct and what is correct is not always obvious.

©ABE and RRC

A sound knowledge and understanding of economics is essential for understanding the business environment and business decision-making. Economics is regarded as a science because it is based on the formal methods of science. It uses abstract models, mathematical techniques and statistical analysis of markets and economies. The aim is to test and apply theories to advance our understanding of both how economies work and the business environment. If you have not studied economics before there is no need to worry if you do not like mathematics, graphs and equations. This *Study Manual* provides an introduction to the study of economics, and its application to business, and maths and equations are kept to a minimum.

Positive and Normative Economics

In the study of economics, because it is a science, an important distinction is made between positive and normative statements. Science is based on theories which are used to make predictions about how some aspect of physical reality works. Successful theories are ones that yield useful predictions and insights into reality. More precisely, successful theories yield predictions that are not refuted when put to the test using real data. Theories that fail to predict correctly are not "good" theories; they are not useful and are unlikely to survive the course of time. Likewise, theories that only predict some things accurately some of the time tend to be replaced or refined. This is how science progresses.

Statements and predictions that can be tested, to see if the theories from which they are derived should be accepted or rejected, are called *positive* statements. *Positive economics* is concerned with such statements: it seeks to understand how economies function by using theories that can be tested in the real world and rejected if they make false predictions. Positive economics is concerned with *"what is"* not with *"what should be"*.

In contrast statements about how the world, or an economy, should be changed to make it better are based on opinions rather than facts. Such statements cannot be proved or disproved using the methods of science. For example, the statement that an increase in the price of petrol will lead to a reduction in the sale of petrol is an example of positive economics. The statement may be right or wrong: the way to find out is to test the prediction using real world data on petrol sales and the price of petrol. On the other hand, the statement that the government should subsidise the price of petrol to help people on low incomes is a *normative* statement. Some people may agree with the statement but others may disagree, because it is based on a value judgement. There is no scientific way of "proving" that it is the correct thing for the government to do. That is, even if we all shared the same values and agreed that the government should help people on low incomes, it does not follow that reducing the price of petrol is the best way to help them. Although this is a simplification, positive economics is concerned with facts while *normative economics* is concerned with opinions.

The Methods of Economic Analysis: the Ceteris Paribus Assumption

The economic behaviour of individuals is complex. The behaviour of consumers and firms interacting in markets is even more complex. The economic decisions and interactions between all the consumers and firms in the economy, with the added complication of actions by the government, make for mind-bending complexity. Economic theory deals with such complexity by using a useful assumption when developing models of economic behaviour, analysing markets and government economic policy. It makes use of the *ceteris paribus* assumption. This is a Latin expression which means *holding other things constant*.

An example is the easiest way to illustrate what it means. Suppose the government of a country has increased the amount of tax it charges on each litre of petrol sold. You have data on the price and the quantity of petrol purchased each day before the tax was increased. You collect data on the quantity of petrol purchased each day following the increase in tax. What your data shows is that the quantity of petrol sold each day has now fallen. Can the fall in the sale of petrol be attributed to the increase in the amount of tax on petrol? It may seem

© ABE and RRC

obvious that the answer is yes. But this would only be a correct inference if it could be shown that none of the other things affecting the demand for petrol had changed at the same time as its price increase due to the government's tax. For example, if the price of cars had been increased at the same time or the price of food had just increased people might have had less to spend on petrol. In other words to study the relation between a change in one factor on another it is necessary to be able to rule out other possible influences operating at the same time. This is where the assumption of ceteris paribus comes in useful. Assuming all other things remain constant, economics is able to demonstrate that for normal goods an increase in their price will lead to a fall in demand.

Microeconomics and Macroeconomics

The functioning of an economy involves the decisions of millions of people as well as the interactions between them. I want to go to town to do some shopping. Should I walk, catch a bus or take my car? If I choose to walk the bus company, the local fuel station and the city centre car park will all be affected: they will have less revenue than if I had decided not to walk to town. Add up all the similar decisions made by thousands or tens of thousands of people a day in just one city, and the revenue implications become significant. If many people decide to switch from using cars to walking or taking a bus because this is better for the environment, then the local fuel station may go out of business and the council and local businesses may suffer a significant fall in revenue. The fuel station closing means unemployment for some people. Reduced council revenue from the car park could mean less support for local amenities. Scale up this example to the entire multitude of decisions taken by all of the people in an economy in a single day, and you can start to appreciate the complexity of the process, and that is just in a day! To make the study of economics more manageable the subject is divided into *microeconomics* and *macroeconomics*.

Microeconomics ("micro" from Greek, meaning small) considers the economic behaviour of individuals in their roles as consumers and workers, and the behaviour of individual firms. It also involves the study of the behaviour of consumers and firms in individual markets. Microeconomic policy includes the different ways in which governments can use taxation, subsidies and other measures to affect the behaviour of consumers and firms in specific markets rather than the economy as a whole. *Macroeconomics* ("macro" again from Greek, meaning large) considers the working of the economy as a whole. It deals with questions relating to the reasons why economies grow, undertake international trade and investment, and experience inflation or unemployment. Macroeconomic policy involves the different fiscal and monetary means through which governments can influence the level of economic activity in an economy. Microeconomics is studied in the first seven units of this subject. Macroeconomics and macroeconomic policy is studied in the remaining units.

A. BASIC ECONOMIC PROBLEMS AND SYSTEMS

Some Fundamental Questions

Economics involves the study of *choice*. The resources of the world, countries and most individuals are limited while wants are unlimited. Economics exists as a distinct area of study because scarcity of resources or income forces consumers, firms and governments to make choices. Economics is concerned with people's efforts to make use of their available resources to maintain and develop their patterns of living according to their perceived needs and aspirations. Throughout the ages people have aspired to different lifestyles with varying degrees of success in achieving them; always they have had to reconcile what they have hoped to do with the constraints imposed by the resources available within their environment. Frequently they have sought to escape from these constraints by modifying that environment or moving to a different one. The restlessness and mobility implied by this conflict between aspiration and constraint has profound social and political consequences

©ABE and RRC

but, as far as possible, in economics we limit ourselves to considering the strictly economic aspects of human society.

It is usual to identify three basic problems which all human groups have to resolve. These are:

- *what*, in terms of goods and/or services, should be produced
- *how* resources should be used in order to produce the desired goods and services
- *for whom* the goods and services should be produced.

These questions of production and distribution are problems because for most human societies the aspirations or wants of people are unlimited. We often seem to want more of everything whereas the resources available are *scarce*. This term has a rather special meaning in economics. When we say that resources are scarce we do not mean necessarily that they are in short supply – though often, of course, they are – but that we cannot make unlimited use of them. In particular when we use (for example) land for one purpose, say as a road, then that land cannot, at the same time, be used for anything else. In this sense, virtually all resources are scarce: for example your time and energy, since you cannot read this study unit and watch a football match – or play football – at the same time.

Choice and Opportunity Cost

Since human wants are unlimited but resources scarce, choices have to be made. If it is not possible to have a school, hospital or housing estate all on the same piece of land, the choice of any one of these involves sacrificing the others. Suppose the community's priorities for these three options are (in order) hospital, housing estate and then school. If it chooses to build the hospital it sacrifices the opportunity for having its next most favoured option – the housing estate. It is therefore logical to say that the housing estate is the *opportunity cost* of using the land for a hospital.

Opportunity cost is one of the most important concepts in economics. It is also one of the most valuable contributions that economists have made to the related disciplines of business management and politics. It is relevant to almost every decision that the human being has to make. Awareness of opportunity cost forces us to take account of what we are sacrificing when we use our available resources for any one particular purpose. This awareness helps us to make the best use of these resources by guiding us to choose those activities, goods and services which we perceive as providing the greatest benefits compared with the opportunities we are sacrificing. This cost will be a recurring theme throughout the course.

You may have been wondering how a community might decide to choose between the hospital, housing estate and school. Which option is chosen depends very much on how the choice is made and whose voices have the most power in the decision-making process. For example, you are probably aware that changing the structure of many of the bodies responsible for allocating resources in the health and hospital services in Britain has led to many strains and disputes. One reason for this was the transfer of decision-making power from senior medical staff to non-medical managers, whose perception of the opportunity costs of the various options available was likely to be very different from that of the medical specialists.

Throughout history societies have experimented with many different forms and structures for decision-making in relation to the allocation of the total resources available to the community. Through much of the twentieth century there has been conflict between the *planned economy* and the *market economy*. In the planned economy decisions are taken mostly by political institutions. In the market economy decisions are taken mainly by individuals and groups operating in markets where they can choose to buy or not to buy the goods and services offered by suppliers, according to their own assessment of the benefits and opportunity costs of the many choices with which they are faced. As the century drew to its

© ABE and RRC

close it was market economies that were in the ascendancy, and this course is concerned mainly with the operation of markets and the market economy. At the same time we need to recognise that market choices have certain limitations and social consequences which cannot be ignored. All the major market economies have important *public sectors* within which choices are made through various kinds of non-market institutions and structures, and economics is able to make a significant contribution to understanding these.

B. NATURE OF PRODUCTION

Economic Goods and Free Goods

The term "goods" is frequently used in a general sense to include services, as long as it does not cause confusion or ambiguity. It is used in this wide sense in this section.

Goods are *economic* if scarce resources have to be used to obtain or modify them so that they are of use, i.e. have utility, for people. They are *free* if they can be enjoyed or used without any sacrifice of resources. A few minutes' reflection will probably convince you that most goods are economic in the sense just outlined. The air we breathe under normal conditions is free, but not when it has to be purified or kept at a constant and bearable pressure in an airliner. Rainwater, when it falls in the open on growing crops, is free, but not when it has to be carried to the crops along irrigation channels or purified to make it safe for humans to drink. Free goods are indeed very precious and people are becoming increasingly aware of the costs of destroying them by their activities, e.g. by polluting the air in the areas where we live.

Production Factors

Since there are very few free goods most have to be modified in some way before they become capable of satisfying a human want. The process of *want satisfaction* can also be termed "*the creation of utility or usefulness*"; it is also what we understand by "*production*". In its widest economic sense, production *includes any human effort directed towards the satisfaction of people's wants*. It can be as simple as picking berries, busking to entertain a theatre queue or washing clothes in a stream, or as involved as manufacturing a jet airliner or performing open heart surgery.

Production is simple when it involves the use of very few scarce resources, but much more involved and complex when it involves a long chain of interrelated activities and a wide range of resources.

We now need to examine the general term "*resources*", or "*economic resources*", more closely. The resources employed in the processes of production are usually called the *factors of production* and, for simplicity, these can be grouped into a few simple classifications. Economists usually identify the following production factors.

- **Land**

  This is used in two senses:

  (a) the *space* occupied to carry out any production process, e.g. space for a factory or office

  (b) the *basic resources* within land, sea or air which can be extracted for productive use, e.g. metal ores, coal and oil.

©ABE and RRC

- **Labour**

  Any mental or physical effort used in a production process. Some economists see labour as the ultimate production factor since nothing happens without the intervention of labour. Even the most advanced computer owes its powers ultimately to some human programmer or group of programmers.

- **Capital**

  This is also used in several senses, and again we can identify two main categories:

  (a) *Real capital* consists of the tools, equipment and human skills employed in production. It can be either *physical capital*, e.g. factory buildings, machines or equipment, or *human capital* – the accumulated skill, knowledge and experience without which physical capital cannot achieve its full productive potential.

  (b) *Financial capital* is the fund of money which, in a modern society, is usually needed to acquire and develop real capital, both physical and human.

Notice how closely related all the production factors are. Most production requires some combination of all the factors. Only labour can function purely on its own, if we ignore the need for space. A singer or storyteller can entertain with voice alone, but will usually give more pleasure with the aid of a musical instrument and is likely to benefit from earlier investment in some kind of training. The hairdresser requires at least a pair of scissors!

Much of economic history is the story of people's success in increasing the quantity and quality of production through the accumulation of human capital and the development of technically advanced physical capital. I can dig a small hole in the ground with my bare hands, but creating the Channel Tunnel between Britain and France has required a vast amount of very advanced physical capital together with a great deal of human skill and knowledge.

Modern firms depend for their survival and success on both their physical and their human resources. While some may feel that the current trend to replace the business term "personnel management" by "human resource management" is in some degree dehumanising, others welcome it as a sign that firms are recognising the importance of employee skills as human capital.

Enterprise as a Production Factor

All economic texts will include land, labour and capital as factors of production. There is not quite such universal agreement over what is often described as the fourth production factor, which is most commonly termed *enterprise*.

The concept of enterprise as a fourth factor was developed by economists who wished to explain the creation and allocation of profit. These economists saw profit as the reward which was earned by the initiator and organiser of an economic activity. This was the person who had the enterprise and special quality needed to identify an unsatisfied economic want, and to combine successfully the other production factors in order to supply the product to satisfy it.

In an age of small business organisations, owned and managed by one person or family, this seemed quite a reasonable explanation. The skilled worker who gives up secure and often well-paid employment to take the risks of starting and running a business is most likely to be showing enterprise. Such a person is prepared to take risks in the hope of achieving profits above the level of his or her previous wage. Many modern firms have been formed in the recent past by initiators, innovators and risk takers of the kind that certainly fit the usual definition of the business entrepreneur. Their names appear constantly in the business press. Few would wish to deny that profit has been and often remains the spur that drives them.

© ABE and RRC

Nevertheless this identification of enterprise in terms of individual risk-taking raises a great many problems when we attempt to apply it generally to the modern business environment. Much contemporary business activity is controlled by very large international and multinational companies such as Microsoft, Toyota, Sony, Philips and Unilever. Who are the entrepreneurs in such organisations? Are they rewarded by profits? How do these companies recruit and foster enterprise? You, yourself, may work in a large organisation. Can you reconcile the traditional economic concept of enterprise as a factor of production with your observations of the structure of your company?

No one doubts the importance of enterprise and profit in modern business. However their traditional explanation in terms of the fourth production factor is at best incomplete and at worst actually dangerous, in that it may be used to justify the very large salaries which company chief executives seem able to award themselves in Britain and the USA.

We shall return to the question of profit in Study Unit 5.

Fixed and Variable Factors of Production

Both economists and accountants make an important distinction between production factors, based on the way they can be varied as the level of production changes. To take a simple example, suppose you own a successful shop. Initially you do not employ anyone but soon find you do not have time to do everything, and are losing sales because you cannot serve more than one customer at a time. So, you employ an assistant. This gives you more time and flexibility and allows you to buy better stock; your monthly sales more than double. You employ another assistant and again your sales increase. You realise, however, that you cannot go on increasing the number of assistants since space in your shop is limited and you can only meet demand in a small local market. You begin to think about opening another shop in another area.

This example helps to illustrate the difference between a production factor which you can vary as the level of production varies, i.e. a *variable factor*, and a factor which you can only move in steps at intervals when production levels change, i.e. the *fixed factor*. In our example the variable factor is the assistants (labour) and the fixed factor is the shop, i.e. land (space) and capital (the shop building and equipment).

In most examples at this level of study it is usual to regard capital as a fixed factor and labour as a variable factor. Although it is not possible to have a fraction of a worker we can think in terms of worker-hours and recognise that many workers are prepared to vary the number of hours worked per week. It is more difficult to have half a shop and even if a shop is rented rather than bought, tenancies are usually for fixed periods. It is more difficult to reduce the amount of fixed factors employed than the variable factors. When a machine or piece of equipment is bought it can only be sold at a considerable financial loss.

This distinction between fixed and variable production factors is very important, particularly when we come to examine production costs in Study Unit 4. It also gives us an important distinction in time. When analysing production, economists distinguish between the *short run* and the *long run*. By short run they mean that period during which at least one production factor, usually capital, is fixed, e.g. one shop, one factory, one passenger coach. By long run they mean that period when it is possible to vary all the factors of production, e.g. increase the number of shops, factories or passenger coaches. Sometimes you may find the short and long run referred to as short and long term. This is not strictly correct, but the difference in meaning is slight and not important at this stage of study.

Production Function

We can now summarise the main implications of our recognition of factors of production. We can say that to produce most goods and services we need some combination of land, capital and labour. At present we can leave out enterprise as this is difficult to quantify. In slightly

©ABE and RRC

more formal language we say that production is a function of land, capital and labour. Using the symbols Q for production, S for land, K for capital and L for labour, (with f for function) this allows us, if we wish, to use the mathematical expression:

$$Q = f(S, K, L)$$

For further simplicity we can use the assumption of *ceteris paribus*, which was explained in the introduction to this unit: we can hold constant the role of two factors of production, land and capital, and concentrate on labour as the only variable input into the production process. That is, as previously noted, we can regard capital and land as fixed and labour as a variable factor.

Total Product

In this section we examine what happens when a firm increases production in the short run, when the firm's available capital and land is fixed and when the only variable factor into the production process is labour. Once again we can take a simple example of a small firm which has a single factory building (land), and a fixed number of machines (capital), installed in its factory. The only way the firm can increase output in the short run is to increase its use of labour. For simplicity we can use the term worker as a unit of labour, but you may wish to regard a worker as a block of worker-hours which can be varied to meet the needs of the business.

Suppose the effect of adding workers to the business is reflected by Table 1.1, where the quantity of production is measured in units and relates to a specific period of time, say, a month. The amount of capital and land employed by the business is fixed. The quantity of production measured here in units produced per month and shown as a graph in Figure 1.1, is, of course, the *total product*. In this example total product continues to rise until the tenth worker is added to the business; this worker is unable to increase total product. This is no reflection on that particular worker who may, in fact, be working very hard. It is simply that, given the fixed amount of capital, no further increase in productive output is possible. The addition of an eleventh worker would actually cause a fall in production. It is not difficult to see why this could happen.

Number of workers	**Quantity of production** *(units per month)*
1	30
2	70
3	120
4	170
5	220
6	260
7	290
8	310
9	320
10	320
11	310

Table 1.1: Number of workers and quantity of production

Suppose the factory has five different machines, each one of which makes a different component for the finished product. Suppose also that each machine is designed to be operated by two workers. When only one worker is employed he or she will have to waste a

© ABE and RRC

lot of time moving between each machine and will not be able to work each machine to its full capacity. Adding a second worker will reduce the time wasted moving between machines and lead to a more than proportional increase in output. As more workers are employed the machines can be progressively operated more efficiently, with two workers to each machine and less and less time wasted by workers moving from one machine to another. As the number of workers employed in the factory increases total product also increases, but at a diminishing rate. Once ten workers are employed then each machine is being operated at its optimum capacity. Adding more workers will not increase production but may actually cause it to fall, as workers start to get in the way of each other and slow the speed of the machines. This is shown in Figure 1.1 by the fall in total product from 320 to 310 when the 11th worker is employed with the fixed number of machines in the factory. Each additional worker's contribution to total product is termed the worker's *marginal product*. Marginal product is the difference in the total product which arises as each additional worker is employed.

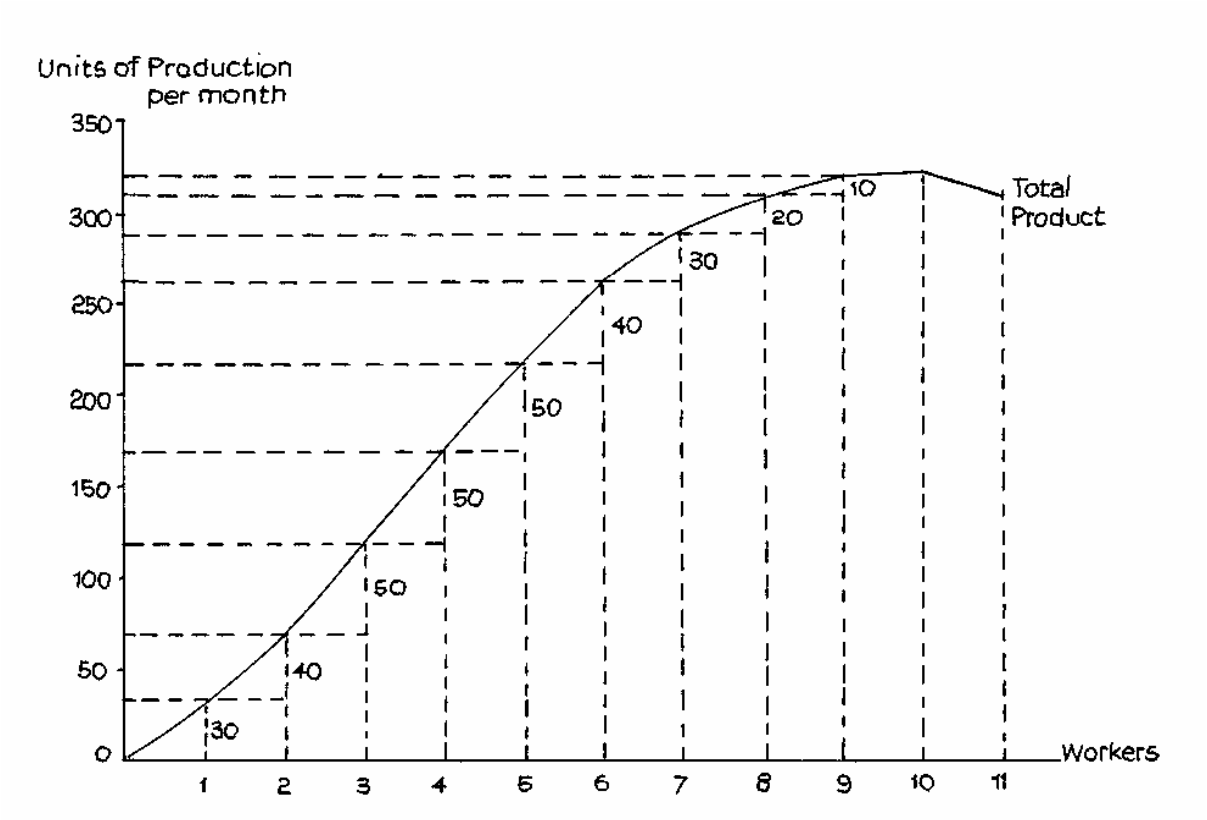


Figure 1.1: Total product

Notice how marginal product changes as total product rises: one worker alone can produce 30 units but another enables the business to increase production by 40 units and one more by 50 units. However, these increases cannot continue and the additional third, fourth and fifth workers all add a constant amount to production. Thereafter, further workers, while still increasing production, do so by diminishing amounts until the tenth worker adds nothing to the total. At this level of labour employment production has reached its maximum, and the eleventh worker actually provides a negative return – total production falls. Perhaps people get in each other's way or cause distraction and confusion. If the business owner wishes to continue to expand production, thought must be given to increasing capital through more machines and, at some point, increasing the size of the factory building to accommodate additional machines and workers. Short-run expansion at this level of capital has to cease. Only by increasing the fixed factors can further growth be achieved.

This example is purely fictional – it is not based on an actual firm; but neither is the pattern of change in marginal product accidental. The figures are chosen deliberately to illustrate some

©ABE and RRC

of the most important principles of economics, the so-called laws of *varying proportions* and *diminishing returns*. It has been constantly observed in all kinds of business activities that when further increments of one variable production factor are added to a fixed quantity of another factor, the additional production achieved is first likely to increase, then remain roughly constant and eventually diminish. It is this third stage that is usually of the greatest importance, this is the stage of *diminishing marginal product*, more commonly known as diminishing returns. Most firms are likely to operate under these conditions and it is during this stage that the most difficult managerial decisions, relating to additional production and the expansion of fixed production factors, have to be taken.

It must not, of course, be assumed that firms will seek to employ people up to the stage of maximum product when the marginal product of labour equals zero, or on the other hand, that they will not take on any extra employees if diminishing returns are being experienced. The production level at which further employment ceases to be profitable depends on several other considerations, including the *value of the marginal product*. This depends on the revenue gained from product sales, and the cost of employing labour, made up of wages, labour taxes and compulsory welfare benefits. The higher the cost of employing labour, the less labour will be employed in the short run and the sooner will employers seek to replace labour by capital in the form of labour-saving equipment.

You should give some thought to the implications of this production relationship for business costs. We will return to it again in Study Unit 4 when we examine costs and the firm's supply curve.

C. PRODUCTION POSSIBILITIES

If individual firms are likely to face a point of maximum production as they reach the limits of their available resources, the same is likely to be true of communities whose total potential product must also be limited by the resources available to the community, and by the level of technology which enables those resources to be put to productive use.

This idea is frequently illustrated by economists through what is usually termed the *production possibilities frontier* (or *curve*), which is illustrated in Figure 1.2.

The frontier represents the limit of what can be produced by a community from its available resources and at its current level of production technology. Because we wish to illustrate this through a simple two-dimensional graph we have to assume there are just two classes of goods. For simplicity, we can call these *consumer goods* (goods and services for personal and household use) and *capital goods* (goods and services for use by production organisations for the production of further goods).

Because resources are scarce in the sense explained earlier in this study unit, we cannot use the same production factors to produce both sets of goods at the same time. If we want more of one set we must sacrifice some of the other set. However, the extent of the sacrifice (i.e. the opportunity cost) of increasing production of each set is unlikely to be constant through each level of production, since some factors are likely to be more efficient at some kinds of production than others. Consequently the shape of the frontier curve can be assumed to reflect the principle of *increasing opportunity costs,* shown in Figure 1.2. In this illustration the opportunity cost measured in the lost opportunity to produce (say) arms is much less at the low level of (say) food production of 2 billion units than at the much higher level of 9 billion units.

The curve illustrates other features of the production system. For example, the community can produce any combination of consumer and capital goods within and on the frontier but cannot produce a combination outside the frontier – say at E. If it produces the mixtures represented by points A, B or C on the frontier all resources (production factors) are fully

© ABE and RRC

employed, i.e. there are no spare or unused resources. The community can produce *within* the frontier, say at D, but at this point some production factors must be unemployed.

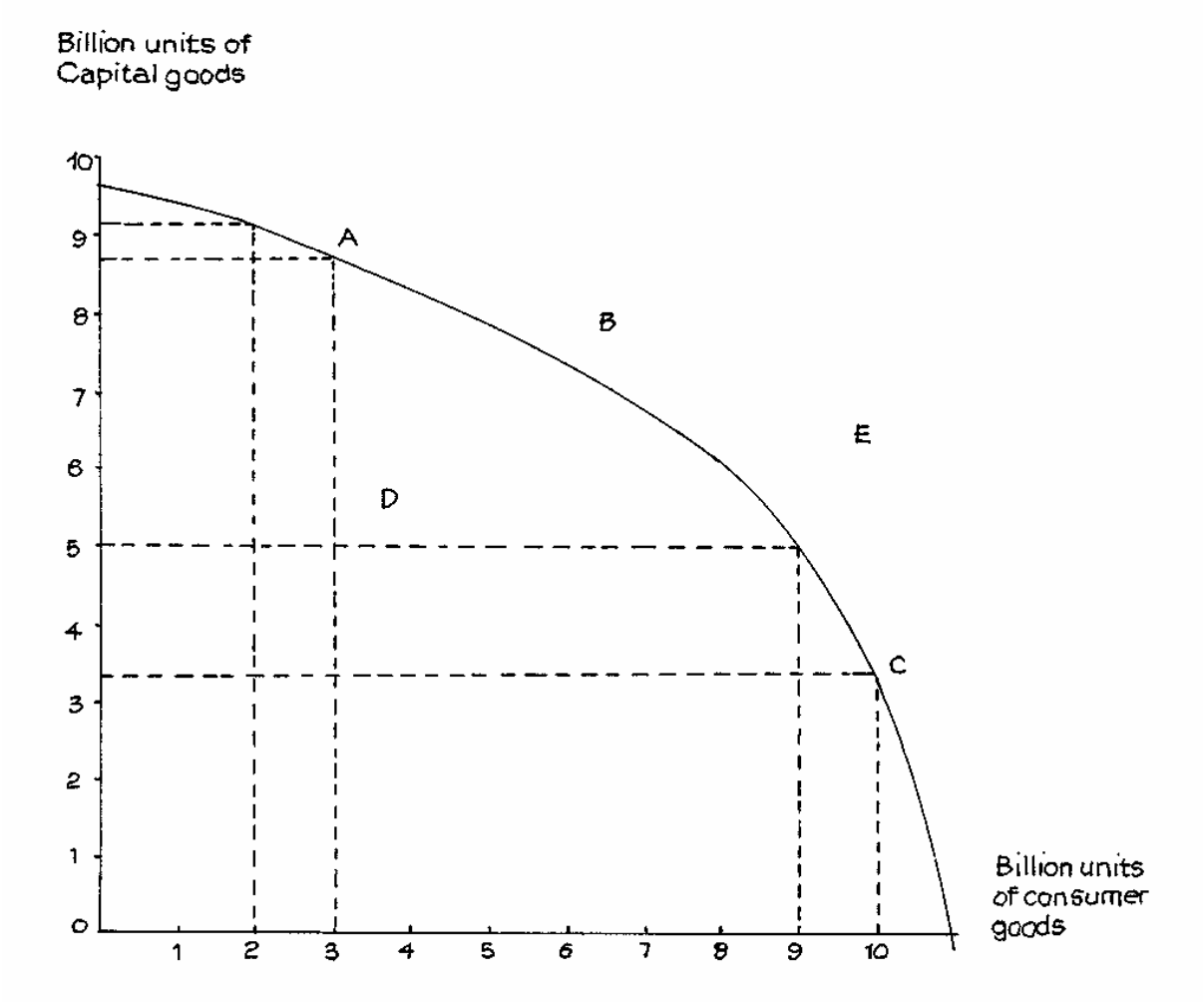


Figure 1.2: The production possibilities frontier

To raise production of consumer goods from 2 to 3 billion units involves sacrificing the possibility of producing 0.3 billion units of capital goods. However when production of consumer goods is 9 billion units, an additional 1 billion units involves the sacrifice of 1.6 billion units of capital goods.

The shape of the curve is based on the principle of increasing opportunity costs.

We can, of course, turn the argument round. If we know that some production factors are unemployed, e.g. if people are out of work, farmland is left uncultivated, factories and offices left empty, then we must be producing within and not on the edge of the frontier. The community is losing the opportunity of increasing its production of goods and services and is thus poorer in real terms than it need be. If, at the same time, some goods and services are in evident inadequate supply – e.g. if there are long hospital waiting lists, many families without homes, some people short of food or unable to obtain the education or training to fit them for modern life – then the production system of the community is clearly not operating efficiently to meet its expressed requirements. Unfortunately it is easier to state these facts than to suggest remedies. There have been very few, if any, examples throughout history of fully efficient production systems where the aspirations of the community have been served by maximum production of the goods and services that the community has desired.

Although generally used in relation to the economy as a whole, the production possibilities (sometimes written as "possibility") curve can also be used to illustrate the options open to a

©ABE and RRC

particular firm. In this case the shape of the curve need not always follow the pattern of Figure 1.2. It might be that if the firm devoted all its resources to the production of one good (in economics the word "good" is used as the singular of "goods") instead of more than one then it would be able to use them more efficiently. They would then gain from what will later be described as increasing returns to scale. In this case the curve would be shaped as in Figure 1.3.

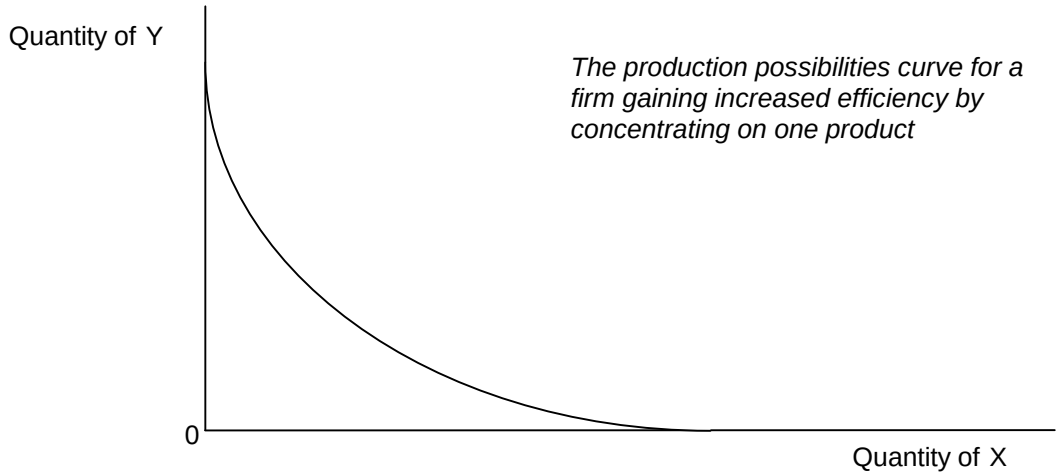


Figure 1.3: Another production possibilities curve

Yet another possibility is that the firm could switch resources without any gain or loss in efficiency, i.e. it would experience constant returns from scale in using its resources. In this case the curve would be linear (a straight line) as in Figure 1.4.

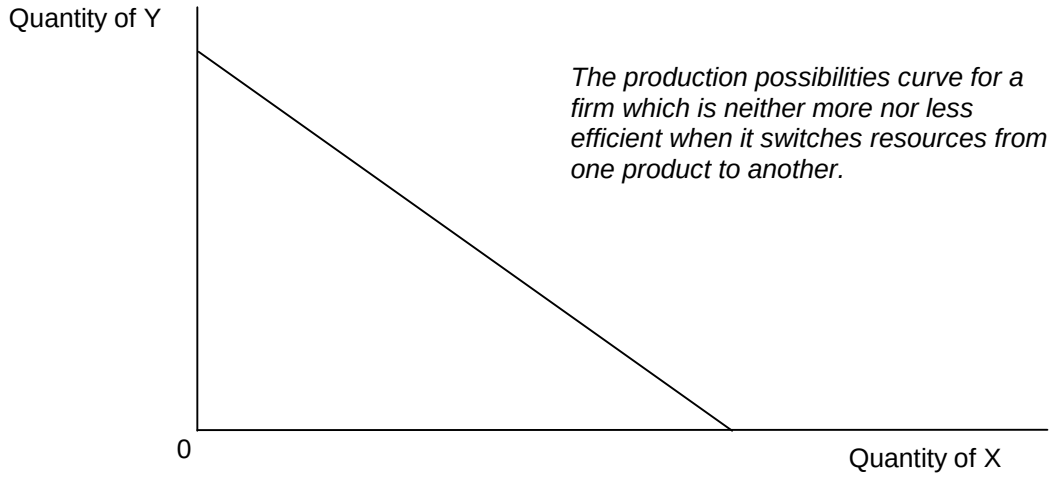


Figure 1.4: A linear production possibilities curve

© ABE and RRC

D. SOME ASSUMPTIONS RELATING TO THE MARKET ECONOMY

Consistency and Rationality

Although we recognise that all people are individuals, and it is usually impossible to predict with complete certainty what actions any individual will take at any given time, nevertheless it is possible to predict with rather more confidence what groups of people are likely to do over a period of time. On this basis it becomes possible to estimate, for example, how much bread will be consumed in a certain town each week or month. A supermarket manager does not know what any shopper will buy when that shopper enters the store, but can estimate how much, on average, the total number of shoppers will spend on any given day in the month. The manager will also know how much is likely to be spent on each of the many classes of goods stocked. Patterns of spending will change of course, but the changes are not likely to be random when applied to large groups. There will be trends that will enable projections to be made into the future with some degree of confidence. As groups, therefore, people tend to be consistent and to behave according to consistent and predictable patterns and trends.

People are also assumed to be rational in their behaviour. Again, we are all capable of the most irrational actions from time to time, but if we behave in a normal manner we are likely to display rational economic behaviour. For example, suppose if given the choice between cornflakes and muesli for breakfast we choose cornflakes, and if given the choice between muesli and porridge we choose muesli. Then, if we are rational, and offered the choice between cornflakes and porridge, we would be expected to choose cornflakes, because we prefer cornflakes to muesli and muesli to porridge. It would be irrational to choose porridge in preference to cornflakes if we have already indicated a preference for muesli over porridge and for cornflakes over muesli.

If we accept consistency and rationality in human behaviour then analysis of that behaviour becomes possible. We can start to identify patterns and trends and measure the extent to which people are likely to react to specific changes in the economic environment, such as price, in ways that we can identify, predict and measure. If we could not do this the entire study of economics would become virtually impossible.

The Forces of Supply and Demand

In studying the modern market economy we assume that the economic community is large and specialised to the extent that we can realistically separate organisations which produce goods and services from those that consume them. We are not studying village subsistence economies which can consume only what they themselves produce. Most of us would have a rather poor standard of living if we had to live on what we could produce ourselves. We can of course be both producer and consumer, but the goods and services we help to produce are sold and we receive money which enables us to buy the things we wish to consume.

As individuals and members of households we are therefore part of the force of consumer demand. As workers and employers we are part of the separate force of production supply. Right at the start of your studies it is important to recognise that supply and demand are two separate forces. These do of course interact (in ways that we examine in later study units) but essentially they exist independently. It is quite possible for demand to exist for goods where there is no supply, and only too common for goods to be supplied when there is no demand, as thousands of failed business people can testify. As students of economics you must never make the mistake of saying that supply influences demand or that demand influences supply.

©ABE and RRC

Basic Objectives of Producers and Consumers

In a market economy we assume that all people wish to maximise their utility. This is simplified to suggest that producers seek to maximise profits, since the object of production for the market is to make a profit and, if given the choice between producing A or B and if A is more profitable than B, we would expect the producer to choose to produce A.

At the same time consumers can be expected to devote their resources, represented by money, to acquiring the goods and services that give them the greatest satisfaction. This is not to say that we all spend our money wisely, or eat the most healthy foods or wear the most sensible clothes. We perceive satisfaction or utility in more complex ways. Economists, as economists, do not pass judgments on the wisdom or folly of particular consumer wants. They recognise that a want exists when it is clear that a significant group of people are prepared to sacrifice their resources to satisfy that want.

When this happens there is demand which can be measured and which becomes part of the total force of consumer demand.

Unfortunately this does not stop some groups of people from seeking to dictate what the rest of the community should or should not want, consume or enjoy. This is a problem of all human societies and is beyond the scope of introductory economics. When Shakespeare's Maria in *Twelfth Night* accused the pompous Malvolio with the damning question "Dost thou think because thou art virtuous there shall be no more cakes and ale?" she was speaking for the market economy in opposition to the planners who would decide for the rest of humanity how to conduct their lives.

Consumer Sovereignty

Although the separation between supply and demand as two different forces has been stressed, the market economy operates on the assumption that, of these forces, consumer demand is dominant. The market production system is demand led: supply adjusts to meet demand. In this sense the consumer is sovereign. Producers who cannot sell their goods at a profit fail and disappear from the production system. Profit is the driving force of the production system: profit is achieved by the ability to produce goods that people will buy at prices that people will pay, while enabling the producer to earn sufficient profit to stay in business – and to wish to stay in business. However strong the demand for goods, if they cannot be produced at a profit they will not, in the long run, be supplied.

If you have lived all your life in a market economy none of this will seem strange to you. But to someone who has lived in a command economy (where production decisions and the quantity, quality and distribution of consumer goods have all been determined by the institutions of the state) the full implications of consumer sovereignty, particularly the implications for individual firms operating in a competitive market environment, can be very hard to grasp.

In the next five study units we shall be very largely concerned with different aspects of the forces of demand and supply and how they interact, or sometimes fail to interact, in the market economy.

© ABE and RRC

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. What is the difference between microeconomics and macroeconomics?
2. How does the assumption of ceteris paribus help in trying to understand economic relationships?
3. Is the following statement an example of a positive or a normative statement?

   "The government should provide free health care for everyone."
4. Is the following statement an example of a positive or a normative statement?

   "When more and more units of a variable production factor are added to a fixed quantity of another factor, the additional production achieved is likely, first, to increase, then to remain roughly constant and eventually to diminish."
5. "For a given size of its budget, the government of a country can only increase its expenditure on education if it reduces its expenditure on roads or defence".

   Which of the following economic concepts is illustrated by this statement?

   (a) normative economics

   (b) opportunity cost

   (c) microeconomics

   (d) marginal product.
6. Can you name a country that has a planned economy? Is your own country a market economy or a mixed economy?

©ABE and RRC

Study Unit 2

Consumption and Demand

Contents ***Page***

A. Utility **18**
Meaning of Utility 18
Total and Marginal Utility 19
Maximising Utility from Available Resources 20

B. The Demand Curve **21**
What is a Demand Curve? 21
Use and Importance of Demand Curves 22
General Form of Demand Curves 23

C. Utility, Price and Consumer Surplus **24**

D. Individual and Market Demand Curves **25**

© ABE and RRC

Objectives

The aim of this unit is to explain the theory of consumer choice using the concept of utility, individual demand and market demand.

When you have completed this study unit you will be able to:

- explain the concept of utility
- explain what is meant by marginal utility, utility maximisation and the property of diminishing marginal utility, using diagrams and/or numerical examples
- explain the relationship between individual utility and individual demand for a good, using examples where required
- solve numerical problems relating to marginal utility and utility maximisation based on utility or consumption data
- identify the difference between individual and market demand.

A. UTILITY

In this unit we introduce the demand curve. The concept of the demand curve is one of the most important concepts used in economics. This is because it provides one of the two keys required to understand how markets work. For this reason it is of great importance for all businessmen and businesswomen. We begin by explaining the concept of utility.

Meaning of Utility

Economists have always faced problems in explaining clearly why people are prepared to make sacrifices to obtain many of the goods and services which they evidently wish to have. In a market economy this difficulty can be stated as "Why do we buy the things we do buy?" Very often we do not "need" them in the strict sense that they are necessary to our survival. In fact our basic needs are really very small, compared with all the things on which we might spend our money in advanced market economies. We can talk in terms of "wants" and recognise that there seems to be no limit to these wants. We also have to recognise that at any given time we are likely to want some things more than others.

What then is the quality that goods must possess that makes us want to acquire them? Clearly this will differ with different goods. Some may be pleasant to eat, some attractive to look at, some warm to wear and so on. The one general term we can apply to all goods and services is that they provide us with *utility*. This does not necessarily mean that they are useful in the sense that they help us to do something we could not do before we had them. It simply means that we perceive in them some quality that makes us willing to make some degree of sacrifice (usually of money) in order to acquire them.

Can we then measure this utility? In an absolute sense, the answer is almost certainly "No". Some economists have proposed adopting a measure called a "util" but no-one, not even the European Commission, has yet proposed that we mark all goods to show how many "utils" they contain. It is more practical to think in terms of money value, since most of us measure the strength of our desire to buy something in terms of the price we are prepared to pay for it. Therefore when an estate agent asks a potential house buyer, "How much are you prepared to offer for this house?" the agent is, in effect, asking the buyer to indicate the value of the utility which the house has for him or her.

More often we find ourselves making comparisons of utility. This arises partly because of the basic economic problem of unlimited wants and scarce resources, so that ranking our wants so we can decide what we can afford to buy is, for most people, an almost daily occurrence. But it also arises because, in modern advanced economies, there is likely to be a range of

© ABE and RRC

different goods to satisfy any particular want. If I want to travel by public transport from Birmingham to Glasgow I could do so by motor coach, by train, or by air. My want is to get from Birmingham to Glasgow, and three options offer the utility to satisfy this want. Each involves different sacrifices of money and time and offers different associated utilities of convenience and comfort. My choice will depend on the resources available to me (how much money I can afford to pay and how much time I have) and on my valuation of the utility afforded by each option. Notice, further, that this utility is not an absolute quality but depends on why I want to make the journey. If it is part of a holiday then I might prefer the coach or train. If I am attending a business meeting from which I hope to achieve a financial benefit and need to be fresh and alert, then the air option is likely to offer the greatest utility – greater, probably, than the price of the fare.

All this may seem very involved, but an appreciation of utility and how it can influence our actions can be a very great help in understanding the true nature of economic demand.

Total and Marginal Utility

Our valuation of the utility provided by any good depends on how strongly we want to acquire it. While there may be several elements involved in this, e.g. we find it attractive or useful, or think it will impress our friends or neighbours, one factor that is always relevant is the amount of that or a similar good we already possess. Suppose I have enough spare cash at the end of the week to buy either a pair of trousers or a pair of shoes but not both, though I would like both. If I already have an adequate supply of trousers for the next few months but do not have any spare shoes then, assuming that their prices are roughly similar, I am likely to buy the shoes. This does not mean that I always value shoes more highly than trousers but that, considering what I already have at the present time, I perceive greater utility in some additional shoes than in additional trousers.

By now, especially if you have remembered the explanation of marginal product in Study Unit 1, you will recognise that I have just given an example of *marginal utility*, i.e. the *change in total utility for a good or group of goods when there is a change in the quantity of those goods already possessed*.

Most of the important decisions relating to the demand for goods and services are influenced by valuations of marginal utility compared with the prices of these goods. The more pairs of trousers I possess the less value am I likely to place on obtaining more, and the more likely I am to spend my available money on other things of comparable price whose marginal utilities are higher.

Willingness to buy thus depends on the comparison of marginal utility with price, and so to some extent it is reasonable to value utility in terms of price. To return to the original house buyer example, if the buyer says to the agent, "My highest offer is £100,000", then for this buyer the value of the marginal utility of the house is £100,000. If this is the buyer's only house then, of course, it is also the total utility.

We must also bear in mind that money itself has utility. Suppose I am saving money for a major holiday or for an expensive durable (long lasting) good such as a house or furniture. Then I may place a high value on money savings and be less inclined to buy trousers and shoes, as long as I have enough of these for my immediate needs. If my income is secure and rising, my valuation of the marginal utility of money could be low and I am more likely to spend it on goods. If my job is not secure and redundancy or retirement is a serious possibility, my valuation of the marginal utility of money is likely to rise, and I will spend less on goods and services. You can easily see the implications of this for the general demand for consumer goods during periods of economic uncertainty, when people think they are likely to have less money in the future. Just as the marginal utility of a good diminishes as the quantity already possessed rises, so marginal utility rises as the quantity of a good already possessed falls – or is expected to fall – in the near future.

© ABE and RRC

Maximising Utility from Available Resources

This relationship between total and marginal utility can be illustrated in a simple graph as in Figure 2.1.

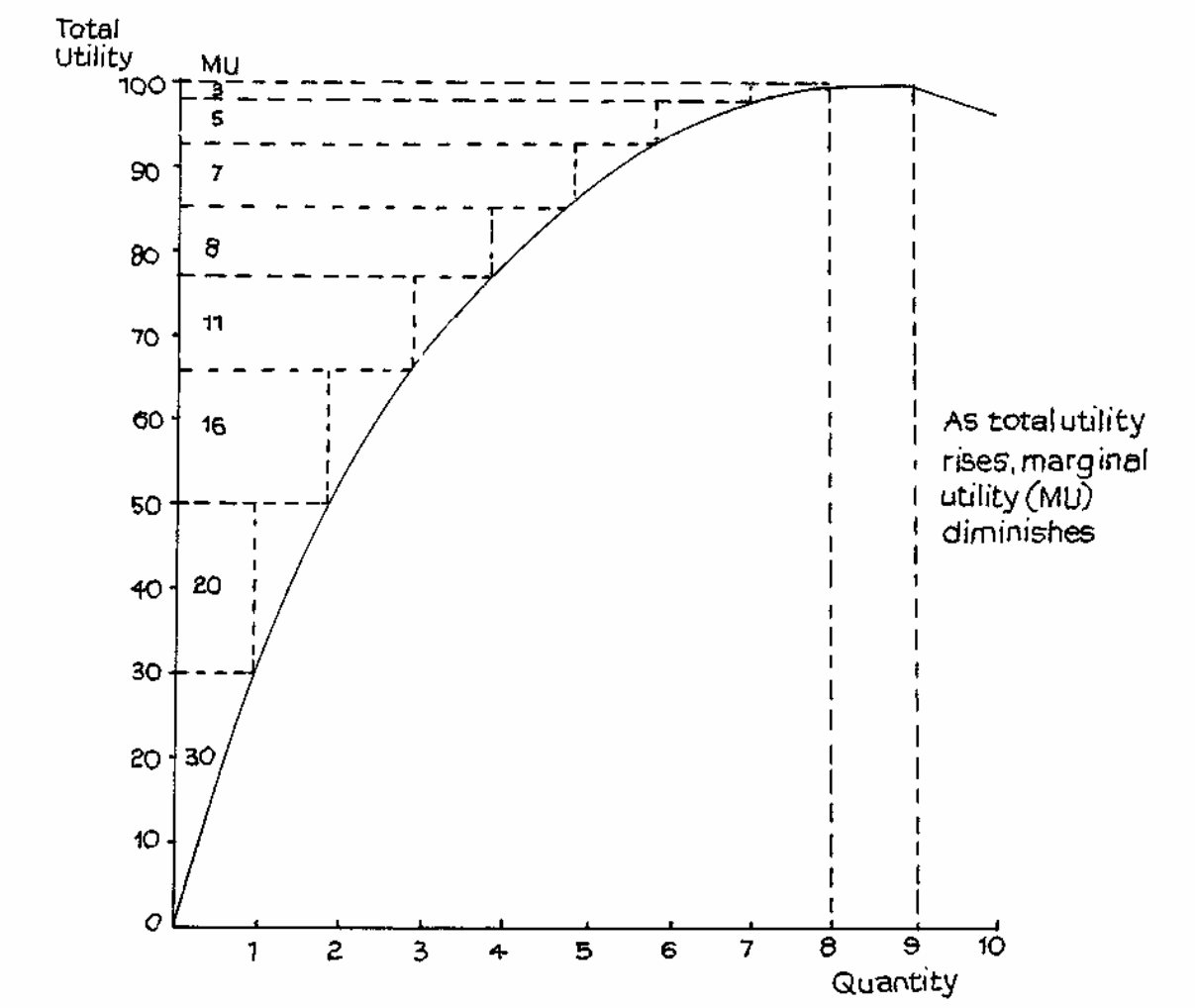


Figure 2.1: Marginal and total utility

Suppose I have no use for more than eight pairs of trousers. This number would provide maximum utility to which we can give a hypothetical numerical value of, say, 100 (representing 100 per cent of the total), but clearly the largest marginal utility would be provided by the first pair. After this purchase the marginal utility of each additional pair diminishes, as indicated by the figures under MU to the right of the vertical axis. The total of 100 is reached with the eighth pair. If I have a ninth, no further utility is added – the total remains at 100. Should I receive a tenth pair my total utility actually falls: perhaps they take up space in my wardrobe I would rather have for something else.

Does this then mean that I should aim at keeping eight pairs of trousers all the time? Not necessarily, since Figure 2.1 takes no account of other important considerations, which include:

- the price of trousers, i.e. the sacrifice I must make to buy them
- my desire for other goods and services, i.e. other marginal utilities (I would not, for example, be too pleased to have eight pairs of trousers if I possessed only one shirt, nor would trousers satisfy my hunger if I did not have enough food to eat)

- how much money I have, i.e. my marginal utility for money.

Only when all these are taken into account would it be possible to estimate how many pairs of trousers would represent, for me, the best total to try and achieve.

Assuming rationality, in the sense explained in Study Unit 1, the most satisfactory quantity of trousers for me would be where my marginal utility gained from the last £1 spent on trousers just equalled the marginal utility per £1 spent on all other available goods and services, and where this also equalled the marginal utility of money. On the assumption that we are valuing utility in monetary terms, the marginal utility of the last £1 of money equals 1.

Putting this statement a little more formally as an equation and using the symbols MU_A to denote the marginal utility for the good A, MU_B for the marginal utility for the good B, P_A for the price of A, P_B for the price of B and so on, we can say that consumers achieve a position of equilibrium in their expenditure when for them:

$$\frac{MU_A}{P_A} = \frac{MU_B}{P_B} = \frac{MU_N}{P_N} = 1 \text{ (which equals the marginal utility of money)}$$

In this state of equilibrium consumers cannot increase their total utility from all goods and services by any kind of redistribution of spending. Spending more on A and less on B, for example, would mean that the marginal utility of A would fall and so be less than that of the marginal utility of B (which would rise) and be less than the marginal utility of other goods, including money. Also the utility gain from A would be less than the utility lost from B so total utility would have fallen. No one rationally spends £1 to receive less than £1's worth of utility.

You may object that this kind of reasoning takes no account of actions such as making contributions to charity, but our use of the term "utility" does embrace such gifts. Presumably we give to a charity because the act of giving to a use we perceive as worthy affords us satisfaction. Therefore it has utility and can be regarded in the same way as other forms of spending. Of course this means, as charities and the organisers of national charitable events have discovered, that giving to charity is also subject to diminishing marginal utility. "Aid fatigue" is the term sometimes used for this.

B. THE DEMAND CURVE

What is a Demand Curve?

So far in this study unit we have considered some of the consequences of price and income changes for the amounts of goods purchased. The general, and in most cases "normal" relationship between price and quantity changes, is frequently illustrated by graphing the anticipated amounts of a good that people can be expected to buy, in a given time period, at a series of different prices within a given price range. This produces a *demand curve*.

Bear in mind that the demand curve is a simple two-dimensional graph. It shows the relationship between just two variables – the price of a good and the quantity of that good that we believe is likely to be purchased over a given time period.

In concentrating on just price and quantity we make the assumption that all other possible influences on demand (quantities of possible purchases) are held constant. These other influences, including income and prices of other goods, will be considered again in the next study unit. For now we can conveniently ignore them. Our concern, for the moment, is with price.

This graph in Figure 2.2 shows the market demand for a good, let's call it X, over the range of prices £12 to £5. That is, it shows how all the consumers in the market for good X vary their weekly purchase of this good as its price rises or falls in the price range £5–£12. It is the market demand curve for the good X.

© ABE and RRC

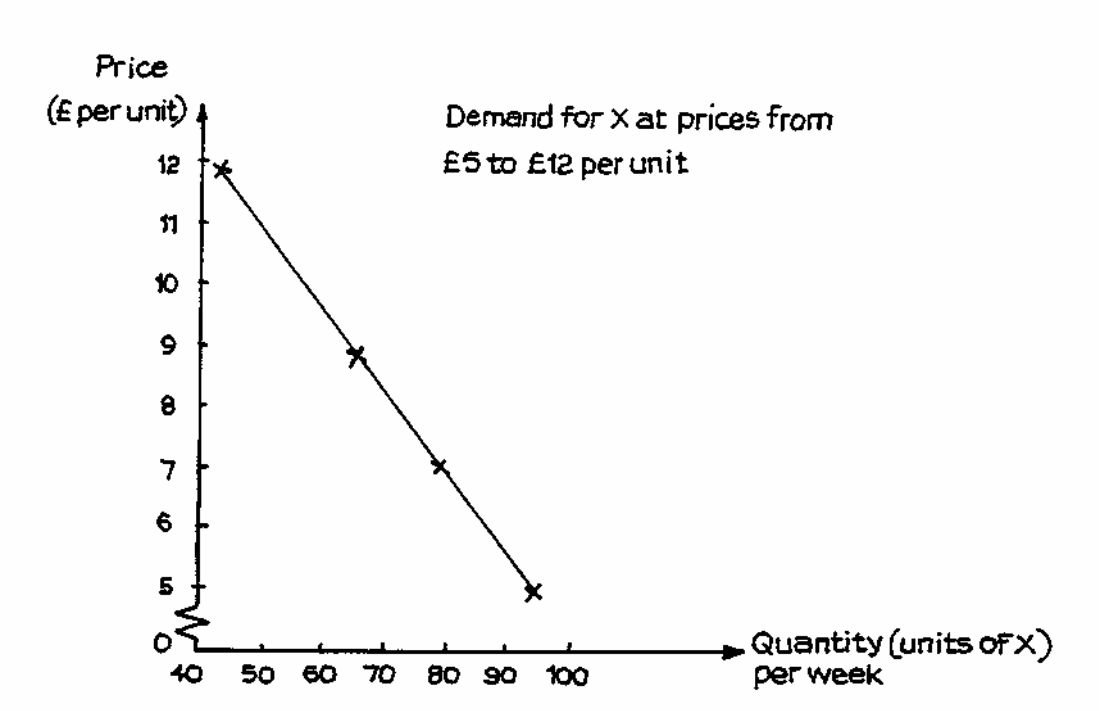


Figure 2.2: A demand curve

This example illustrates the general shape of the demand curve and the normal relationship between price and quantity demanded of a product. If all other influences remain constant, we would expect the quantity demanded to rise as price falls and to fall as price rises. Notice that, in our example, we have made the following assumptions:

(a) The price of all other goods and services remains constant as the price of good X changes. That is, we are making use of the simplifying ceteris paribus assumption once again.

(b) The incomes of consumers also remain constant when the price of good X changes.

(c) Another point to remember is that we are considering here a flow of demand related to a set period of time. It is always necessary to do this. We cannot compare a weekly amount at one price directly with a monthly amount at another. When we change one variable – here price – to analyse its effect on quantity, we have to keep all other elements constant, including the time period to which the stated quantity relates. In our example, this period was a week.

Use and Importance of Demand Curves

As you will see as you progress through this course, the demand curve is used extensively in economic analysis. The *price-quantity relationship* is one of the most important things we need to know when considering sales of products. A firm must know the likely result of a change in price, because any alteration in quantity demanded will affect the total sales revenue.

Governments also need to know the probable effects of any change in a tax imposed on products. Because such a tax will influence price, the price-quantity relationship is again an important issue. If a government is considering an increase in a tax such as value added tax, which influences a very wide range of goods, it needs to know what extra total revenue it can expect to gain from the tax increase. It cannot assume that quantities consumed of all goods affected will remain the same; it must take into account the probable changes in quantity demanded that will result from the changes in price.

© ABE and RRC

General Form of Demand Curves

At this stage of study, you will meet demand curves chiefly in relation to general analytical problems. Actual figures are then less important than the general shape and slope of the curves. It is therefore normal to draw general curves, in which price and quantity are denoted simply by letters. For reasons that will become clearer in later study units, it is simpler to draw what are called "linear curves" (i.e. straight-line graphs) for part only of the full price and quantity range. This is because, for most purposes, we are concerned only with a limited range of possible prices and quantities. When there are special reasons for departing from these normal practices, we shall explain them. Examples of typical general demand curves are given in Figures 2.3 and 2.4.

Notice that in Figure 2.3 a given change in price appears to produce a greater change in quantity demanded than in Figure 2.4. This assumes that both figures are drawn to the same scale. You must remember that the steepness of a demand curve will be affected by the scale of the (horizontal) X-axis, and graphs must be drawn to the same scale, so that comparisons can be made.

It is a convention or general rule in economics that price per unit is measured on the vertical axis or Y-axis, while quantity in units per period of time is measured along the horizontal axis X-axis. It is often customary to label the axes simply "Price" and "Quantity".

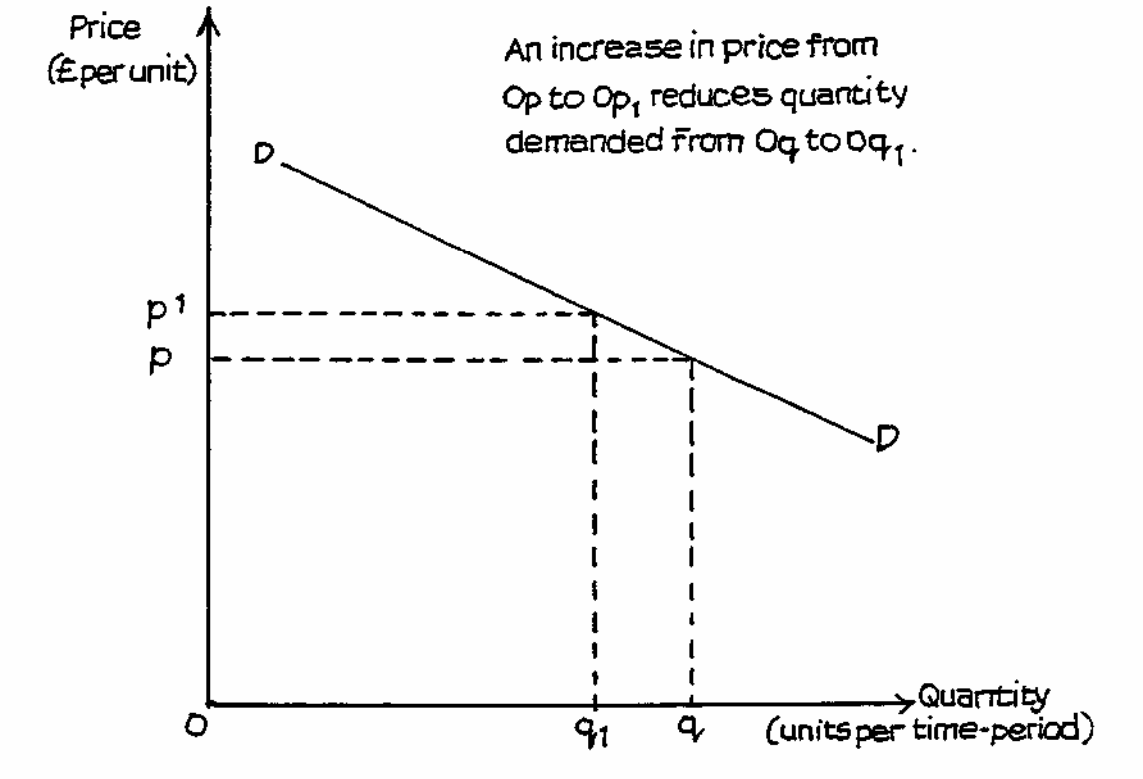


Figure 2.3: General demand curve

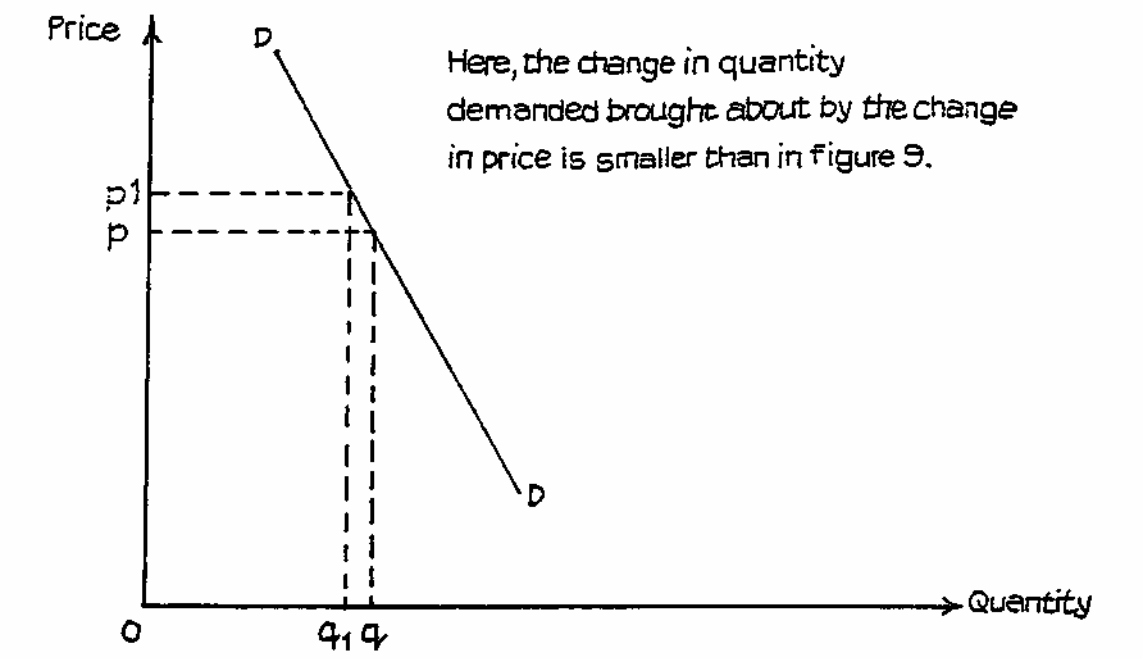


Figure 2.4: Another demand curve

C. UTILITY, PRICE AND CONSUMER SURPLUS

The idea of utility is not too hard to grasp. We recognise that we will only buy something if (for us) it satisfies a want. In other words, if it is of some use to use: for us it possesses utility. We can also appreciate that the utility we perceive for one more unit of a good depends on how much of that good we already have. Suppose I have some apple trees in my garden. In a year when, for some reason, the trees bear very little fruit, I value highly the few apples that do grow and will go to some trouble to pick them carefully when they are ripe. However, in another year the same trees may fruit abundantly and produce more apples than I really want. In that year I may not bother to pick them all, and may allow some to stay on the trees or lie on the ground. Thus, to me, the value of the apples depends on the quantity available and is equal to their *marginal utility* – the usefulness to me of some additional apples to those I already have.

The same principle applies if I have no trees at all and I have to buy apples or any other goods. I will only pay the price to obtain them if this price is not more than the value of their marginal utility. This idea gives us a means of putting a monetary value on marginal utility. Let us say that I like to eat apples but do not have to do so; other fruit readily is available. I will only buy them at a price I consider reasonable. Suppose that, in a particular week, I see that apples are priced at 160p per kilo. This to me is dear, and above my valuation of the utility of a kilo of apples. I do not buy any. Next week the price has fallen to 120p per kilo, but I still think this is too dear and again I do not buy. The third week the price has fallen to 100p per kilo. I give this more thought but, in the end, still do not buy. By the fourth week, the price has fallen to 80p per kilo, and this time I am prepared to buy a kilo. My marginal utility for apples is such that 80p is the highest price I am prepared to pay for a kilo of apples. I can thus put a value on my marginal utility for a kilo of apples: it is 80p.

Suppose now that the next time I visit the store the price of apples has fallen yet again and it is now 60p. Again I buy a kilo. The value of my marginal utility for a kilo of apples has remained at 80p and I would have been prepared to pay 80p, but the price asked by the store was only 60p, so this is what I paid. Consequently I gained a surplus of 20p. The value

© ABE and RRC

of my sacrifice was less than the value of the additional utility I gained: the difference was a surplus to me.

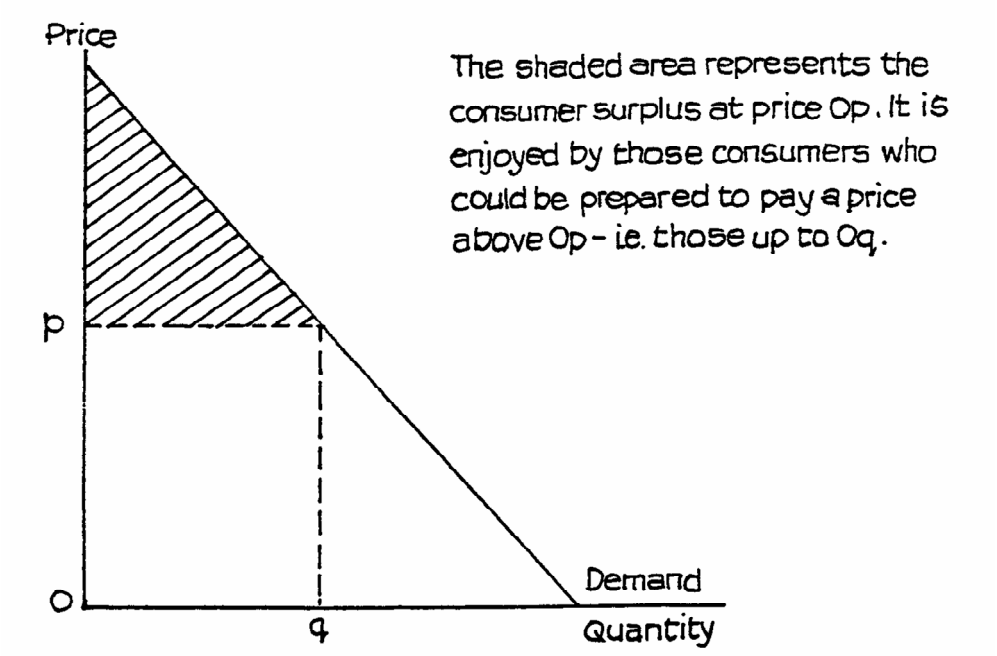


Figure 2.5: Demand curve – consumer surplus

Since the price of 60p per kilo was below my valuation of the marginal utility of a kilo of apples I might decide to buy two or perhaps three kilos. In this case I was valuing the marginal utility of the additional amount bought above my usual quantity at less than the 80p but still now below 60p. If, as seems likely, most consumers react in this way, then we have no difficulty in accepting the general shape of the demand curve outlined in the previous section: that is people are prepared to buy more of a good at a lower than at a higher price.

These ideas are illustrated in Figure 2.5, which shows a normal demand curve for a product the price of which is "p" on the graph. The fact that the demand curve extends to prices higher than p indicates that there are consumers who are willing to pay a higher price. However, if the price charged is p, then these consumers achieve a surplus which is represented by the shaded area.

The demand curve is downward sloping to indicate that more of the product will be bought as the price falls. This follows the assumption that most people will buy more of a product if they think the price is favourable. Marginal utility diminishes as the quantity already possessed rises. So, to sell more, the supplier is likely to have to reduce price. Remember that, as always, when considering the effect of one change we make the assumption that other things remain unchanged. In practice they will not, and in the next study unit we recognise this. My valuation of the marginal utility of apples will change if I discover that the store has received a large consignment of nectarines and peaches and is selling these at prices around my marginal utility for these fruits.

D. INDIVIDUAL AND MARKET DEMAND CURVES

Although we do not think in these terms every individual has their own *individual demand* curve for each of the goods and services they are interested in consuming. How do we know this? Because ask any person how much they would like to buy of something at a particular price and you will get an answer! Knowledge of an individual's demand curve is required to answer questions relating to how a particular individual is likely to react to the change in the

© ABE and RRC

price of a good or service. However, for many purposes what interests economists, firms and governments is not how a specific individual will respond to a change in the price of a good (say because the government has put a tax on the price of the good), but how all consumers in the market for the good respond to the change in its price.

For example, suppose a firm making bottled fruit juice drinks is faced with an increase in costs due to an increase in the price of fresh oranges. How much will the firm's weekly sales of its bottled orange drink fall if it passes on its increase in costs and puts up the price of its orange drink? To answer this question the firm needs to know what the *market demand* curve for bottled orange drinks looks like.

The market demand curve for a good or service is the *horizontal summation* of all the separate individual demand curves for the good or service. What this means is that the quantity demanded at different prices by each person is combined with the quantity demanded by all the others in the market, to give the total quantity demanded at each and every price. This is illustrated in Figure 2.6 for a simplified market with only two customers.

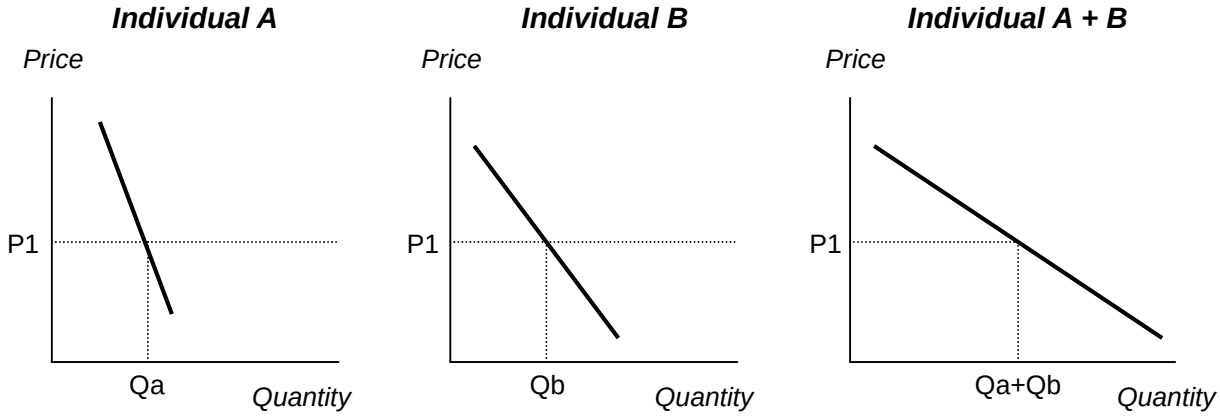


Figure 2.6: Demand curve illustrating horizontal summation

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. Why would a person who likes chocolate, who has just consumed five bars, be unwilling to pay as much for a sixth bar of chocolate as they did for the first bar?
2. What is consumer surplus?
3. What factors are assumed constant when constructing an individual's demand curve for a good?
4. What information would you need to have to construct the market demand curve for a good?

Study Unit 3

Demand and Revenue

Contents ***Page***

A. Influences on Demand **29**
Flow of Demand 29
Product's Own Price 29
Prices of Other Products 30
Income Available for Spending 30
Price and Availability of Money and Credit 30
Market Size 30
Advertising or Marketing Effort 30
Taste 31
Expectations 31
Special Influences 31
Summary of Influences 31
The Relative Importance of Influences 31
Shifts in the Demand Curve 32
Some Further Considerations 32

B. Price Elasticity of Demand **33**
Calculation 33
Influences on Price Elasticity of Demand 35

C. Further Demand Elasticities **36**
Income Elasticity of Demand 36
Influences on Income Elasticity of Demand 37
Cross Elasticity of Demand 37
Influences on Cross Elasticity of Demand 37
The Importance of Elasticity Calculations 38

D. The Classification of Goods and Services **38**
Normal Goods 39
Inferior Goods 39
Giffen Goods 39
Luxury Goods 40
Bads 40

(Continued over)

© ABE and RRC

Substitutes 40
Complements 40

E. Revenue and Revenue Changes 41
Total Revenue 41
Average Revenue 42
Marginal Revenue 44

© ABE and RRC

Objectives

The aim of this unit is to: explain the concept of elasticity in relation to different types of good and firm behaviour through an understanding of the revenue function; solve numerical problems involving elasticity.

When you have completed this study unit you will be able to:

- explain the reasons for movements along or shifts in demand curves
- identify the formulae for, and explain what is meant by, own-price, cross-price, and income elasticities of demand and discuss factors which affect each of these elasticities
- solve numerical demand elasticity problems using demand information
- explain, in words, diagrams and with reference to demand elasticities, what is meant by each of the following: normal goods, bads, inferior goods, Giffen goods, luxury goods, complements and substitutes
- identify real world examples of each of these
- examine, using diagrams and numerical examples, the relationship between total revenue, average revenue and marginal revenue and between marginal revenue and the elasticity of demand for a profit- maximising firm
- discuss how a profit-maximising firm might respond to information about demand elasticities.

A. INFLUENCES ON DEMAND

Flow of Demand

The demand curve which we identified in Study Unit 2 illustrates the quantities of a product that a group of consumers are prepared to buy at a range of possible prices. We must remember that these quantities are always related to a time period. Demand is seen in terms of a flow of purchases over a stated time. For example a greengrocer may want to know the weekly quantity of apples he can sell at a price of 80p per kilo, and compare this with the weekly quantity he could sell at 90p per kilo. The time is not always shown in simple demand graphs, but we must not forget its importance. It is not much use being able to sell 100 kilos instead of 50 kilos if it takes three times as long to do so.

If we clearly understand this idea of demand flow, remembering the points we made in Study Unit 2, we can go on to identify the various influences which affect that flow.

Product's Own Price

This is regarded as the most important influence on demand: normally, we expect a rise in price to lead to a fall in quantity demanded, and a price fall to produce a rise in quantity.

Therefore in general we can accept that, if all other considerations are equal (which they seldom are), people will prefer to pay a lower price rather than a higher price for a product the quality of which they know and accept.

We should also recognise that expectations of future price movements can influence current demand. If people expect prices to rise next week, they will if possible prefer to buy now at the lower price. On the other hand, this may be regarded as a temporary distortion of demand which will have little effect over a longer period of time. If the longer-term effect is not taken into account, it might look as though demand was rising as prices rose – when in fact people had taken the view that a price rise today was likely to be followed by further rises tomorrow, and were acting accordingly.

© ABE and RRC

If a new product is introduced to the market, there is likely to be an effect on other goods. For instance the introduction of cheap electronic calculators destroyed the demand for slide rules. On the other hand the development of portable radios and personal stereos also created a demand for the associated (complementary) product – the batteries needed for their operation. If a major product is introduced and becomes popular enough to absorb a significant part of personal income, then people will reduce purchases of other products which they may consider less desirable. There may be no obvious association between the desired product and the one neglected. For example, a person who decides to pay for a part-time degree course to enhance career prospects may think it worthwhile to spend less on entertainment or to put off replacing a car or furniture.

Prices of Other Products

Sometimes, two products are clearly associated – petrol and motor oil or motor car tyres, for instance. A rise in petrol costs may lead to a fall in the use of cars and hence to reductions in demand for oil and tyres. Even when products are not directly linked, a change in the price of one may still influence a wider demand. If a man smokes heavily and is unable to check his habit, a rise in tobacco prices will lead him to spend less on a wide range of other products. In the same way, a rise in mortgage interest will force families to spend less on other goods.

Income Available for Spending

For the majority of goods and services, i.e. for *normal* goods, we would expect the change in demand to be in the same direction as the change in income. But for some *inferior* goods, the changes would be in the reverse direction, so that a rise in income produces a fall in demand and vice versa.

Notice that a good is inferior only if it is perceived as offering less satisfaction for a particular type of want. Thus, as a normal means of transport a motorcycle may be perceived as inferior to a car even though, as a piece of engineering, it may be superior. Suppliers may be able to revive demand for an inferior good by changing its appeal; adapted and marketed as a sporting and leisure good the motorcycle has enjoyed such a demand revival, and as such is often bought by people who also possess cars.

Price and Availability of Money and Credit

Many goods are bought with the help of borrowed money (credit). Money and credit have an influence on demand separate from the effect of income. If the cost of credit (i.e. the rate of interest) rises there is likely to be a reduction in demand for the more expensive goods.

Market Size

Many factors can change market size. A firm selling clothes to teenagers will benefit from any increase in the numbers of teenagers in the population. Specialist shops selling babies' and children's wear will suffer from a declining birth rate. Market size can be increased by improvements in communications and technology. The development of the Internet has greatly increased the market area open to many consumer-goods firms. Increased foreign travel by people from a country can extend the demand and market area for foreign wines and foods in that country. Improved techniques of refrigeration extended the market for frozen vegetables.

Advertising or Marketing Effort

Very few products sell themselves. Most have to be marketed, and the more extensive the advertising effort, the more that is likely to be sold. Some marketing specialists suggest that there is a direct relationship between a firm's share of market advertising and its share of

© ABE and RRC

market sales. Certainly, it is the volume of advertising in relation to competitors' advertising that is likely to be important.

Taste

This is a quality difficult to define. People's desire to buy products is the result of many influences, not all of which are fully understood. Fashions change, and these changes cannot always be caused by advertising. The successful firm is often the one that is able to make an accurate prediction of changes in fashion and taste.

Expectations

Expectations of future changes in any of the previously mentioned influences can affect present demand. For example, people expecting rising prices will buy now rather than later. On the other hand, if they fear unemployment and falling incomes, they will cut down their present spending. Notice that these reactions may actually help to bring about the feared future changes.

Special Influences

Certain products may be subject to special influences other than the ones we have already mentioned. The demand for soft drinks or for waterproof clothing, for instance, will be influenced by weather conditions. The demand for private education in an area will be influenced by the reputation of State-owned schools in that area.

Summary of Influences

All these influences on demand for a product can be expressed in a form of mathematical shorthand. Thus, we can say that:

$$Q = f(P_o, P_a, Y_d, N, A, T).$$

This simply means that the quantity demanded of any product (Q) is a function (f) of (is dependent upon) its own price (P_o), the prices of other goods (P_a), disposable income (Y_d), market size (N), marketing effort (A), and customer taste (T).

The Relative Importance of Influences

Of course the relative importance of these influences varies for different products, and it is necessary for suppliers to estimate this if they are to avoid damaging errors. For example, if price is not of first importance, a price reduction will simply reduce revenue and profit. In such a case perhaps the supplier would have more to gain from increases in price and advertising expenditure.

Suppliers can attempt to estimate the relative importance of the demand influences by recording and measuring the effect of those, such as price and advertising, under their control, and also noting the effects of other measurable changes such as movements in average incomes. Much information may also be gained from market research, e.g. by asking people why they favour certain brands and what their reactions would be to price movements. In some cases, shopping simulations can be staged with people given a certain amount of money and then asked to spend it on a range of goods displayed in a store. The scientific study and calculation of demand functions from information gained from all available sources is known as *econometrics*. In some cases these studies have resulted in calculations that have proved remarkably accurate, but in other cases have been less successful. There are many things that can go wrong in the estimation of future demand! Business decisions still have to be made against a background of market uncertainty.

© ABE and RRC

Shifts in the Demand Curve

A normal two-dimensional graph can cope with only one influence in addition to quantity changes. For this reason, because the normal demand curve relates quantity to the product's own price, a change in quantity demanded brought about by a change in one or more of the other influences must be represented graphically by a shift in the whole demand curve.

Suppose there is an increase in disposable income which increases the quantity demanded at each price within a given range. This effect can be shown as in Figure 3.1, where the price remains constant at Op but the increase in income has shifted the curve from DD to D_1D_1, so that the quantity demanded at Op rises from q to q_1. A fall in income or a decline in taste for example would produce the reverse result, i.e. a shift from D_1D_1 to DD.

Remember always to distinguish a movement along a demand curve produced by a change in price (all other influences remaining unchanged) as shown in Figure 3.1 from a shift in the whole curve, showing that demand has moved at all prices within the range under consideration.

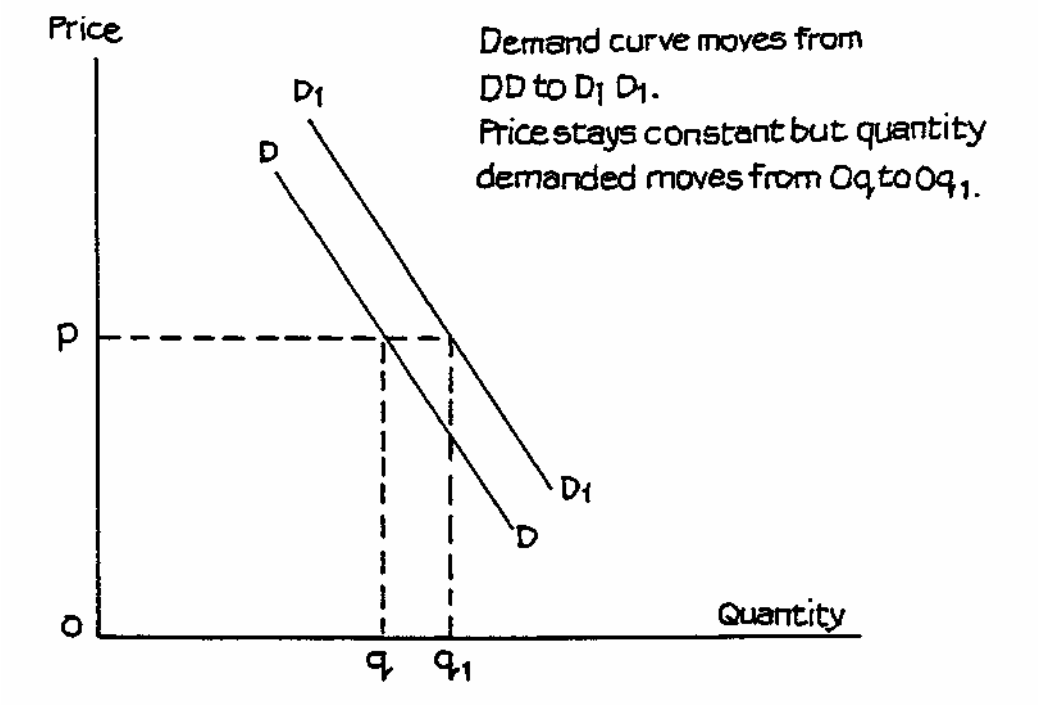


Figure 3.1: A shift in the demand curve

Some Further Considerations

It has been argued that the "normal" influences we have identified do not tell the full story, and that a fuller understanding of social psychology can give further insights into consumer behaviour. For example, supermarket chains are well aware of the importance of impulse buying, when goods are skilfully displayed. There is also a recognised "snob" effect, when goods may be bought because they are expensive and they appear to be indicators of the owner's wealth and status. While these considerations are interesting and are clearly of importance to marketing specialists, we can include them under the more general headings of advertising and taste, for the purposes of general analysis of consumer demand.

© ABE and RRC

B. PRICE ELASTICITY OF DEMAND

We have now seen that there is a definite relationship between price and quantity changes. This is most important for practical studies of price and sales movements, because it determines how sales revenue responds to changes in selling price. We need to have a precise way of measuring and analysing the various possible relationships between demand, price and sales revenue. Because demand is seen as stretching and shrinking in response to price movements, the concept we use is called the *price elasticity of demand*.

Calculation

Price elasticity of demand can be denoted by the symbol E_d. It is the relationship between a proportional change in quantity demanded and a proportional change in price, such that:

E_d = proportional change in quantity demanded ÷ proportional change in price, or

$$E_d = \frac{\Delta Q}{Q} \div \frac{\Delta P}{P}$$

where:

P = price of the product

Q = quantity demanded of the product

ΔQ = a small change in Q and

ΔP = a small change in P.

As explained earlier, for the great majority of goods a rise in price leads to a reduction in quantity demanded and a fall in price leads to an increase in quantity demanded. Thus the change in quantity is the reverse of the change in price. One of the changes will be negative, indicating a reduction: thus the value of E_d will also be negative. In some older text books this used to be ignored but the general tendency today – and the one you should follow – is to keep strictly to using this negative sign. So:

- When the calculation of price elasticity of demand produces a result which is more negative than –1, i.e. when the proportional change in quantity is greater than the proportional change in price, we say that demand is *price elastic*.
- When the calculation of price elasticity of demand produces a result which is less negative than −1, i.e. when the proportional change in quantity is less than the proportional change in price, we say that demand is *price inelastic*.
- When the calculation of price elasticity of demand produces a figure of –1, i.e. when the proportional change in quantity is equal to the proportional change in price, we say that demand has *unitary elasticity*.

The demand for fish is likely to have a price elasticity of around –0.9, that for washing powder about –0.3, and that for eggs around –0.02. These demand elasticities are price inelastic but fish is clearly much more price sensitive than eggs. Note that while the demand for washing powder is price inelastic, for a particular brand of washing powder it might well be price elastic, say around –1.3.

One important feature of price elasticity of demand is that it changes as price changes. Consider the demand curve shown in Figure 3.2.

At point A, E_d = –1, so demand is neither elastic nor inelastic. Here, revenue remains the same at both prices because the change in price produces exactly the same proportional change, so the size of the ratio $\Delta Q/Q$ is the same as the size of the ratio of $\Delta P/P$.

© ABE and RRC

At point B, E_d is more negative than –1, so that demand is price elastic. This means that the size of the ratio of $\Delta Q/Q$ is greater than the size of the ratio of $\Delta P/P$.

A reduction in price at B results in a more than proportional increase in quantity demanded, so that there is an increase in total revenue. A firm in this position will increase revenue by reducing price but lose revenue if it increases price.

At point C, the position is completely reversed and E_d is less negative than –1, so that demand is price inelastic. This means that the size of the ratio $\Delta Q/Q$ is less than the size of the ratio of $\Delta P/P$.

A reduction in price here results in a less than proportional increase in quantity demanded, so that there is a fall in total revenue. A firm in this position will lose revenue by reducing price but gain revenue by increasing price.

The point of greatest possible revenue on any linear demand curve is where price elasticity is at unity (where $E_d = -1$).

Notice also that the calculations shown in this illustration are made around the midpoint of each change. Calculations made in this way are called "arc elasticity". They are the correct way to measure price elasticity, unless we are able to use the necessary mathematical techniques to calculate "point elasticity" at a particular point on the demand curve. For all but very small changes, point elasticity calculations will show different results depending on whether we assume a price rise or a price fall, and this is confusing and inaccurate. You can test this for yourself if you compare the calculation for a price rise from £9.50 to £10.50 with a price fall from £10.50 to £9.50.

© ABE and RRC

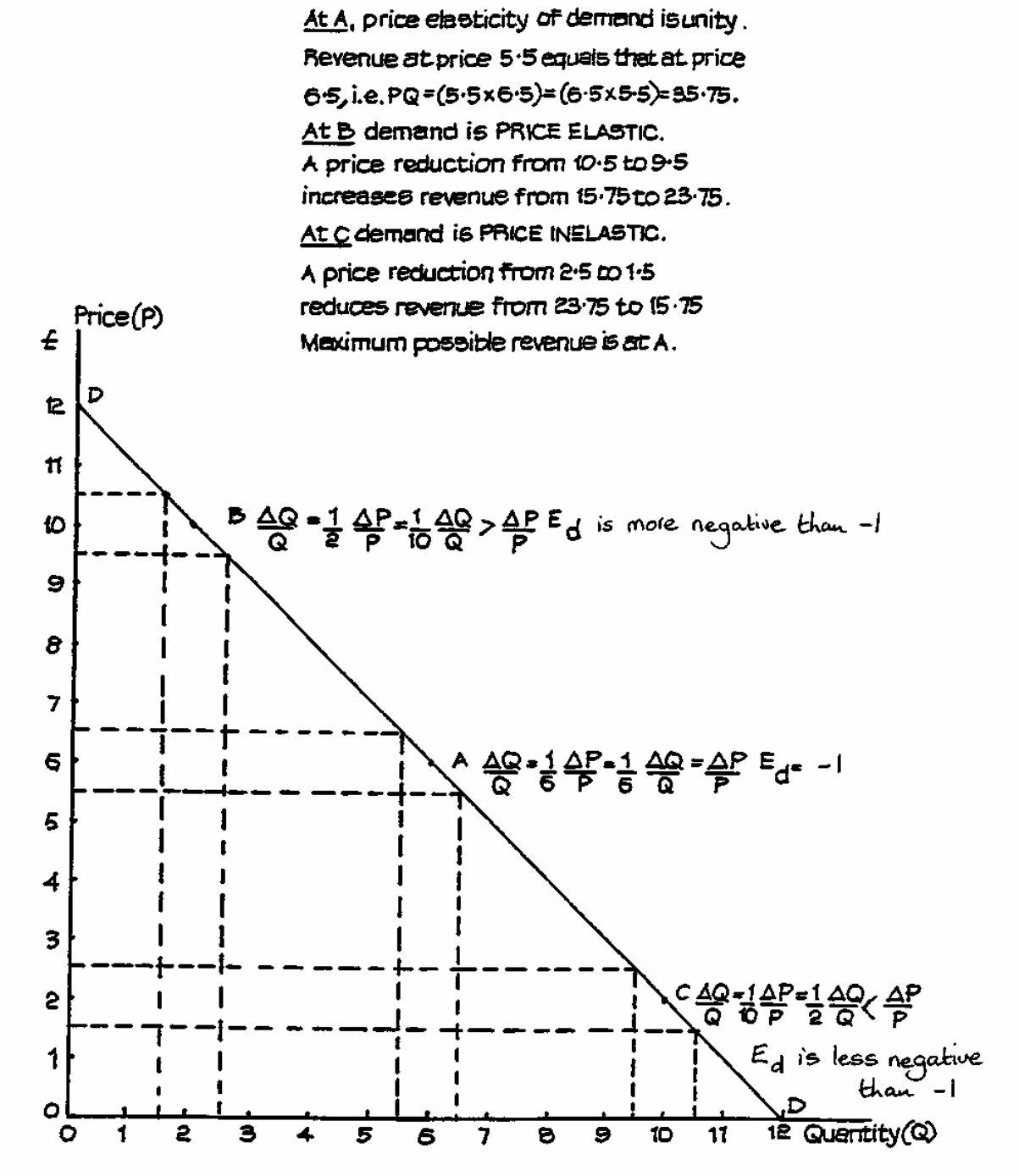


Figure 3.2: Change in price elasticity as demand changes

Influences on Price Elasticity of Demand

We have seen that the price elasticity of demand can be expected to change as price changes, so that the product's own price can normally be regarded as an influence on its elasticity. The important point is whether buyers are likely to pay much attention to the price when deciding whether to buy, or if other influences are more important. These influences may include current fashion or social attitudes, strong habits (even addiction, in some cases such as tobacco smoking) or the need to buy in order to achieve some other desired objective, such as buying petrol in order to drive to work.

© ABE and RRC

If the product price is only a relatively small amount compared with normal income, then price is likely to be less important than the other influences affecting demand, which is thus likely to be price inelastic. Toothbrushes, matches, and shoe polish are all examples of products likely to be price inelastic. Here, high relative price changes at normal price levels are unlikely to weigh heavily with consumers, because annual spending on these items is only a very small part of total income. Other influences, e.g. social attitudes (toothbrushes), smoking decline, the move away from coal fires (matches), and development of non-leather shoes (polish), are likely to be much more important.

We must also be careful to distinguish between the demand elasticity for the class of product and that for a particular brand of the product. My decision whether or not to buy household soap is not likely to be greatly influenced by a 10 per cent rise in its price. However when I am actually making my purchase, I am quite likely to compare the prices of two brands and choose the cheaper, assuming that I do not think one is superior in quality to the other. Thus, demand for a product can be price inelastic, whereas demand for a specific brand of the product can be price elastic. This difference can often be seen in foods. Families may keep to a tradition of the Sunday joint of meat and pay roughly the same price for this each week, thus showing a demand price elasticity of around unity (i.e. 1). However, the choice of which meat to buy can be very much influenced by its price, so that we can expect the demand price elasticity for pork, beef and lamb, and certainly for some particular cuts of beef and lamb, to be higher than unity, especially if the general level of all meat prices has been rising.

C. FURTHER DEMAND ELASTICITIES

The general concept of elasticity can be applied to any of the influences on demand. If you think about the concept, you will realise that it is simply the ratio of a proportional change in quantity demanded to the proportional change in the influence considered to be responsible for that change in quantity. The only limiting element in using elasticity is that the influence must be capable of some sort of precise measurement or evaluation. This makes it difficult to produce a definite calculation for changes in taste or fashion for instance, as this is very difficult to measure. The most commonly used elasticities, in addition to the product's own price, are those for income and for other prices.

Income Elasticity of Demand

Income elasticity of demand relates to proportional change in quantity demanded to the proportional change in disposable income of customers for the product.

It can be denoted by E_y, so that

> E_y = proportional change in quantity demanded ÷ proportional change in disposable income.

This may be positive or negative, because there may be an increase in demand following an income increase or a fall in demand. If the income and quantity changes are in the same direction, then the figure for E_y is positive. If the changes are in the opposite directions to each other, then the figure carries the negative sign (–). A rise in income usually leads to a rise in demand, but demand for some goods may fall. In many countries in recent years the demand for bicycles has fallen as incomes rise and people switched to cars. Such goods are known as *inferior goods*.

Notice that we are referring here to "disposable income", i.e. the income left to the consumer after compulsory deductions have been taken. The most important of these deductions are income tax and National Insurance contributions. We may also include contributions to pension schemes or to trade unions or professional bodies, where membership is necessary for employment.

© ABE and RRC

In recent years, some economists have argued that we should really be thinking in terms of "discretionary income". This is the income that is left from disposable income after all the regular and largely essential household payments, over which the individual has very little control, have been made. The deductions which would be made to arrive at discretionary income would be such items as rent or mortgage interest repayments, water and sewerage charges, essential fuel charges (gas and/or electricity) and possibly the cost of travelling to and from work. When these items have all been allowed for, the amount of discretionary income (the income that people are genuinely free to spend as they choose) is usually very small in relation to the original gross income.

Influences on Income Elasticity of Demand

The following influences are likely to increase a product's income elasticity of demand:

- A high price in relation to income. If a period of saving is required before purchase is possible, or if consumers have to borrow money to obtain a product, then demand can increase only when an income rise makes this possible.
- If goods are preferred to "inferior" substitutes, then people may be ready to buy more of these when income increases make this possible.
- Association with a higher living standard than that currently enjoyed is likely to lead to rising demand when incomes do rise.

In general, the more highly-priced durable goods (household machines, motor vehicles, etc.) and services are more likely to be income elastic than the staple items of food and clothing. We do not usually buy twice as much of these if we receive double our former income. On the other hand, our spending on holidays may increase by far more than double. Increased spending on motor transport is also associated with rising incomes. Although we have been considering income rises, very similar comments apply to income reductions. Holidays and motor cars are often the first things to be sacrificed in the face of a sudden drop in income.

Cross Elasticity of Demand

Cross elasticity of demand relates the proportional change in demand of one product to the proportional change in price of another:

E_x = proportional change in quantity demanded of X ÷ proportional change in price of Y.

Again, the demand movement may be in the same or the opposite direction to the price movement, and the same rules for negative signs apply.

If two products are substitutes for each other, we can expect a rise in price of one to lead to a rise in demand for the other. Beef and pork are in this position, or meat and fish.

However if the two products are linked together, e.g. petrol and motor car tyres, then a rise in price in one leads to a fall in demand for the other, and E_x carries the negative sign (–).

Influences on Cross Elasticity of Demand

The more close substitutes a product has, the more likely it is to react to changes in price of any of those substitutes. The demand for coach travel reacts to changes in rail fares. In the UK the link became closer when motorways cut down the times of road journeys between the major cities, and long-distance coaches became more directly comparable with intercity trains. Brands of goods are normally much more cross elastic with each other than the good itself is with other goods. We are not unduly influenced by other price movements when we decide how much soap to buy, but we are much more ready to switch to a competing brand when there is a rise in the price of the brand we normally buy.

© ABE and RRC

In the same way, the intensity of negative cross elasticity depends on how closely products are associated with each other. For people in England, the demand for suntan lotion is likely to rise if the price of air travel and holidays in the sun falls.

The Importance of Elasticity Calculations

The calculation of elasticities is not just of academic interest. Anyone who wishes to predict accurately the effect of changes in price or income on revenue and on quantities bought needs to have a clear idea of elasticity and its calculation.

If a business manager thinks that a price rise will always increase sales revenue, then he or she needs to be reminded that this is far from being true. A price rise when demand is price elastic will, as you have seen, reduce total sales revenue.

Governments making changes in income or expenditure taxes must be able to calculate their effects on demand. If they do not, then their predictions about the results of the tax change are likely to prove badly out of line with reality.

A government wishing to increase its tax revenue will tend to choose goods for which the demand is price inelastic – tobacco for example, or petrol. However if it goes on increasing the tax, the time will eventually come when demand becomes price elastic. Any further increase will result in a reduction in sales revenue and a fall in tax receipts. This can be seen by referring to Figure 3.2, where a price rise from 5 to 7 (for example) will move the good to that part of the demand curve where price rises produce a reduction in total revenue.

Price elasticity of demand can also change as a result of other influences. If, for example, there is a long-term trend away from smoking, we can expect demand for cigarettes to become price elastic at lower price levels in the future.

If governments wish to influence consumer demand by price changes, they are likely to try to make demand more price elastic by ensuring that suitable substitutes are available for the target product. For instance, to reduce consumption of leaded petrol, the availability and demand for unleaded petrol must be encouraged, and vehicle engines must be capable of easy and cheap conversion to unleaded petrol. They may wish to support any tax changes by changes in the law, perhaps requiring all new vehicles to be adapted to use unleaded fuel.

D. THE CLASSIFICATION OF GOODS AND SERVICES

In this section we provide a summary of what we have said concerning elasticities. We do this by examining how the properties of demand curves and the different measures of elasticity can be used to classify goods and services, in ways that are helpful when analysing market situations for firms' pricing decisions and in product development and marketing strategies.

In economics goods can be classified as being:

- normal goods
- inferior goods
- Giffen goods
- luxury goods
- bads
- substitutes
- complements.

© ABE and RRC

Normal Goods

The vast majority of goods and services in the world are normal goods. The demand curve for normal goods slopes downwards from left to right. As explained previously, the defining characteristic of a normal good is that it has a positive income elasticity of demand. A luxury good is a special case of a normal good in that it is a good with a positive and high income elasticity of demand. As incomes increase the demand curves for normal goods shift outwards to the right as shown in Figure 3.1.

Inferior Goods

The demand curve for an inferior good also slopes downwards from left to right. The defining characteristic of an inferior good is that it has a negative income elasticity of demand. As incomes increase the demand curves for normal goods shifts inwards to the left, indicating that less is demanded at each price. In contrast, a reduction in incomes will shift the demand curve for an inferior good to the right.

Giffen Goods

Giffen goods (named after named after Sir Robert Giffen, who is attributed as first suggesting the existence of such goods) are a special case of inferior goods. A person's demand for inferior goods decreases, ceteris paribus, as their income increases and increases as their income decreases. That is, as we have said, inferior goods have a negative income elasticity of demand. For people on very low incomes their demand for a good may actually increase as the price of the good increases. This is because the rise in price reduces their *real income* to such an extent that they cannot afford to buy sufficient of more preferred goods. Real income refers to the quantity of goods and services a person can buy with their money income. If I have £300 a week to spend and the prices of all the things I buy each week double, my real income falls because I can now only buy half the quantity with my £300.

The demand curve for Giffen goods slopes upwards from left to right, unlike the demand curve for normal goods, with more demanded at a higher price than at a lower price. The negative real income effect associated with the rise in price outweighs the desire to buy less because of the higher price.

In practice Giffen goods are rare. Examples are the types of food items that form an important part of the daily diet of people on very low incomes. Potatoes, bananas or rice, as a source of carbohydrate, are the main daily foods for many of the world's most impoverished people. Depending on their tastes, and their incomes, they may supplement their consumption of one of these sources of carbohydrate with some meat or fish and/or vegetables. But if the price of potatoes rises significantly, some people may be so poor that they can no longer afford to buy potatoes, and fish and vegetables. Faced with a choice between feeling hungry because they can only afford very, very small amounts of potatoes, fish and vegetable on their plate or feeling full because of a large plate of potatoes, they may buy more potatoes despite their higher price.

Strictly the term "Giffen" applies only when the "inferior" income effect created by a change in price is more powerful than the normal price substitution effect which leads people to switch their expenditure in favour of goods as they become relatively cheaper. However it is often used more widely whenever demand appears to rise as price rises for whatever reason. There are a number of other possible explanations for this behaviour. For example, people may (rightly or wrongly) associate price with quality, e.g. for tomatoes, and prefer to pay a little more in anticipation of obtaining a more satisfactory fruit. If there were some other trusted mark of quality, the normal price-quantity relationship would hold. Demand may also rise for a work of art which people think is gaining acceptance in the art world. If people think that the price is going to rise even more in the future, they may buy the work of art as an investment and not simply because they get pleasure from looking at it. In this case, we are

© ABE and RRC

really dealing with a different product. In yet more cases, the rise in demand is just the result of other influences as described in this study unit, and these are proving more powerful than the influence of price on its own.

Luxury Goods

Luxury goods are usually high-priced goods, often with a well-known brand name. In marked contrast to Giffen goods, the income elasticity for luxury goods is positive, as it is for normal goods. As people's real incomes increase we observe that the pattern of their demand changes: they start to buy goods that they did not purchase when their incomes were low. The demand curve for luxury goods is downward sloping, as for normal goods, but the whole demand curve shifts outwards to the right as consumers' real incomes increase. This rightward shift of the demand curve for luxury goods is very pronounced. This is because in the case of luxury goods the income elasticity of demand is not just positive but it is greater than one. If a person had an income elasticity of demand for a particular good of say +3, this would imply that their demand for the good would increase by 300 per cent if their income doubled.

Although the demand curve for some goods that appear to be luxury goods can be upward sloping, like that for a Giffen good, meaning that demand increases as price rises, the economic reason for this is different to that for Giffen goods. In fact, it is better to call these goods "*snob*" goods, to indicate that they are a special case of luxury goods. The demand for snob goods increases as their price increases for the reason that people attach importance to their price as a desirable, possibly the most desirable, characteristic of owning and using the good. Does a £10,000 bottle of wine taste that much better than a similar wine costing £100? The answer does not matter for some people: they are buying the £10,000 bottle of wine as a statement or display of their wealth, and the very high price is the thing that shows this! You should be able to think of similar examples involving some makes of luxury car, watches, trainers and ladies' fashion.

Bads

"Bads" are simply those things that we would rather not have but which may nevertheless exist, and be consumed in the sense that people have no choice but experience them. Examples include atmospheric pollution, water pollution, noise and crime. By definition there are no demand curves for bads, at least for most people. The concept is still useful, however, because it explains why communities and governments may take action to intervene in markets to reduce or eliminate the production of certain goods and services that are associated with the production of bads, e.g. manufacturing equipment which causes a high level of pollution.

Substitutes

As explained in an earlier section, when the relationship between the demand for one good and the price of another, as measured by their *cross-price elasticity of demand*, is *positive* the two goods are referred to as *substitutes.* That is, an increase in the price of one of the two goods will lead to an increase in the demand for the other. Conversely, a decrease in the price of one will lead to a decrease in demand for the other. For example, a decrease in the price of digital cameras will lead to a decrease in the demand for traditional film-based cameras.

Complements

As explained earlier, when the relationship between the demand for one good and the price of another, as measured by their *cross-price elasticity of demand*, is *negative* the two goods are referred to as *complements*. That is, an increase in the price of one will lead to a

© ABE and RRC

decrease in the demand for the other. For example, a large increase in the price of cars will lead to a decrease in the demand for petrol.

E. REVENUE AND REVENUE CHANGES

We have seen that there is a definite relationship between price and quantity changes. This is most important for practical studies of price and sales movements. We now need to study the different concepts of revenue used in economics and the relationship of revenue and the elasticity of demand.

Total Revenue

In general revenue refers to the money received from the sales of a product. For this reason, the term "sales revenue" is often used. To have any practical meaning, revenue should also be related either to a time period or to a definite quantity of goods sold. For example, a shopkeeper may refer to her weekly sales revenue (the total amounts of sales achieved in a week) or to her revenue from the sales of n pairs of shoes or k kilos of potatoes. A statement that her revenue is £y means nothing, unless we can relate it to some quantity of time.

Revenue will not always increase as more goods are sold – this will be the case only if a firm can continue to charge the same price, regardless of quantity it sells. If I make leather belts and can sell all the belts I can make at a standard price of £5, then my total revenue is always £5 multiplied by whatever quantity I sell.

This can be shown in the form of a total revenue curve, as in Figure 3.3.

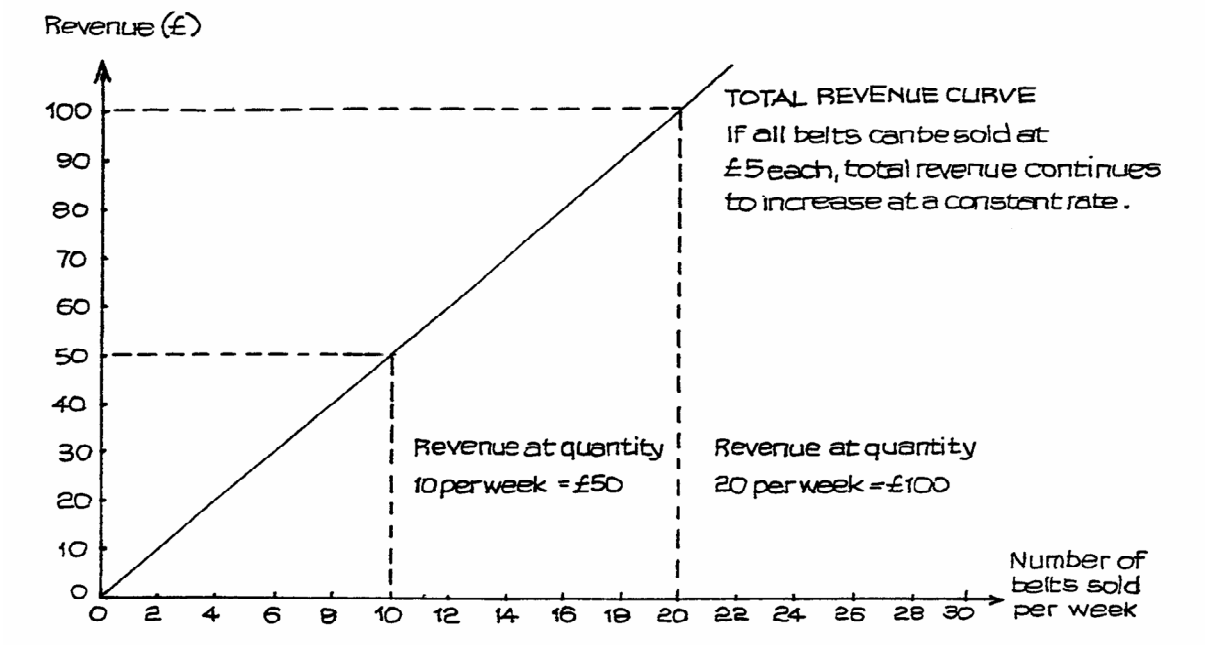


Figure 3.3: Total revenue curve

However, if I continue to produce more and more belts, there will come a time when customer resistance sets in. I shall have difficulty in finding more people who value belts at this price of £5, i.e. the marginal utility of which is at least £5. When this time comes, I may still find more people who are willing to pay £4.

Now, in developed market economies shopping conditions are such that shoppers expect all goods to be priced, so I cannot leave my belts without any price ticket attached and hope to sort out from the people who visit my shop those willing to pay £5 and those willing to pay

© ABE and RRC

£4. If I want to sell more belts and am willing to charge £4, then I must charge this price to everyone. If I continue to produce even more, I might then find that to sell the increased quantity I have to charge £3. If I go on doing this, I am likely to find that my total revenue starts to fall.

Suppose I find that total revenue rises if I reduce the price from £5 to £4, but falls if I reduce the price to £3. This will happen if the reduction from £4 to £3 does not produce enough additional sales to make good the loss suffered when I charge £3 to those people who would still have bought at prices of £5 or £4. My sales schedule at the three prices might be as in Table 3.1.

Price per belt	Number of belts I can sell per month	Total revenue
£5	200	£1,000
£4	280	£1,120
£3	340	£1,020

Table 3.1: Sales schedule for belts

This effect can be shown in the form of a simple graph but this time the turning point can be seen (Figure 3.4). If I try to reduce the price still further, below £3, I shall lose even more revenue.

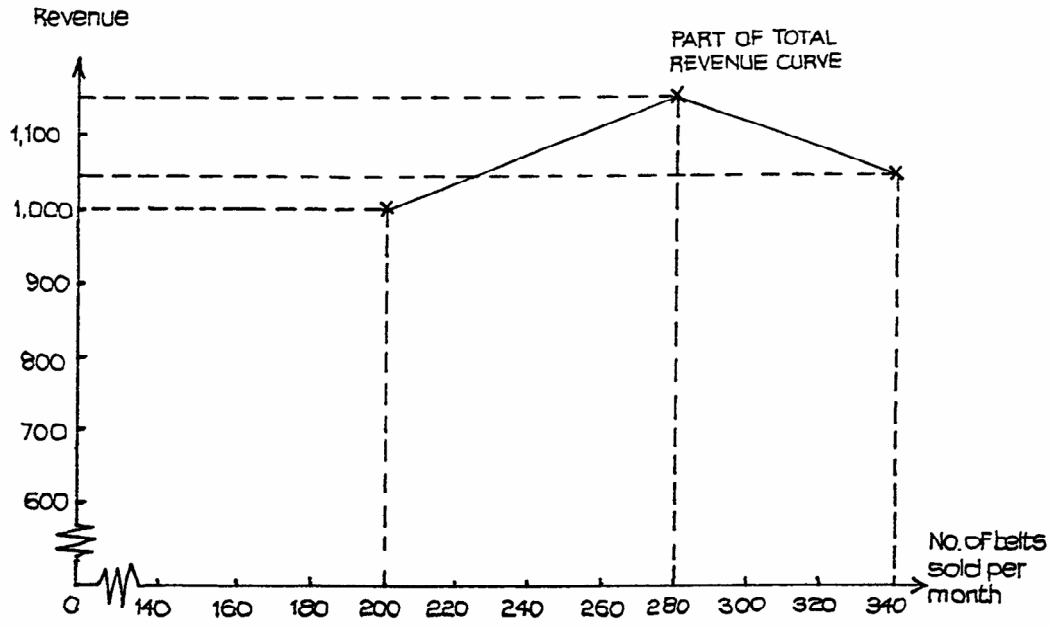


Figure 3.4: Revenue from sale of belts

Average Revenue

We are going to use the term "average" in its most common sense: the average revenue is the total revenue divided by the quantity of goods sold. If a shop's weekly revenue from selling broccoli is £600 and it sells 300 kilos in the week, the average revenue of the broccoli sold is £2 per kilo.

© ABE and RRC

If all goods are sold at the same price in the given time period – as, say, with our leather belts – then the average revenue is the same as the price. The average revenue curve for the belts is shown in Figure 3.5.

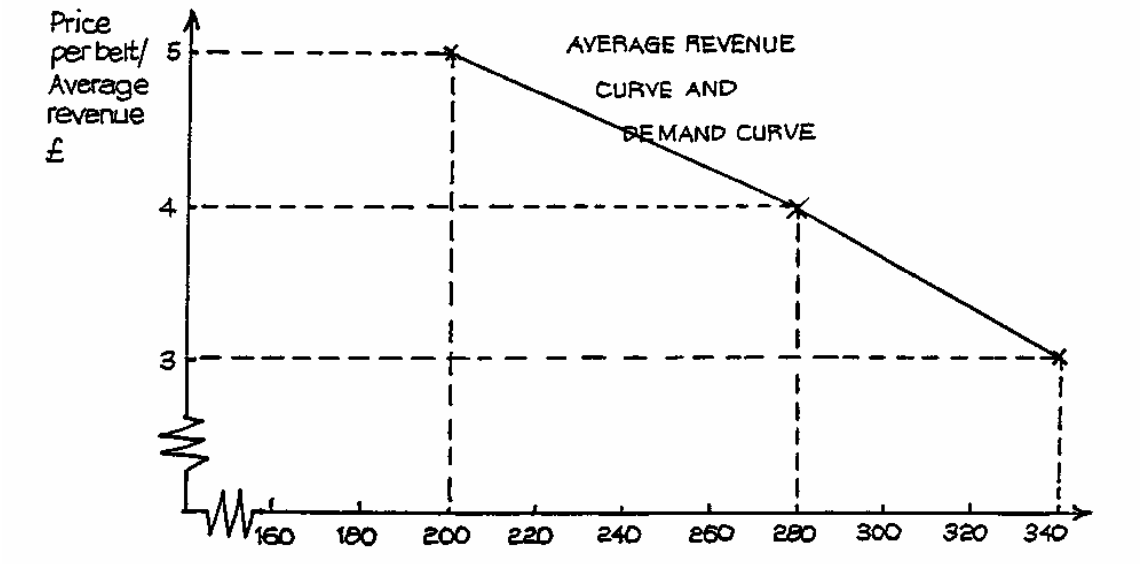


Figure 3.5: Average revenue curve for belts

Notice in this case that the average revenue curve is really just the same as the demand curve. This will always be the case where all items sold in the time period are sold at the same price, i.e. where there is no price discrimination between different customers.

In most market conditions a firm's average revenue curve is identical with its demand curve and the two terms can be used interchangeably.

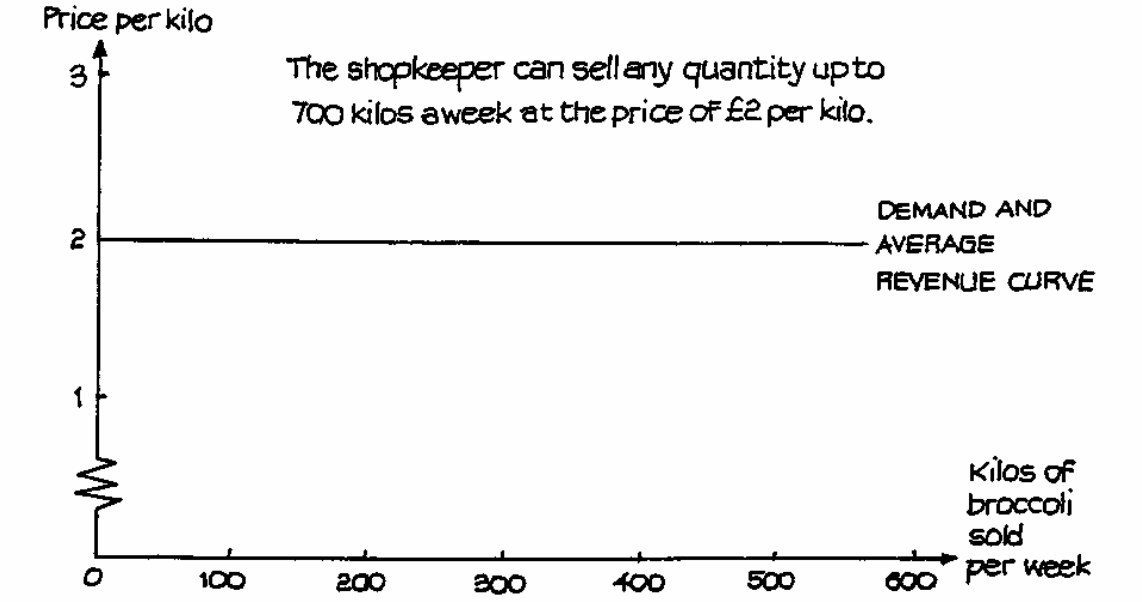


Figure 3.6: Horizontal average revenue curve

To return to our shopkeeper selling broccoli at £2 per kilo: let us suppose that she is selling every kilo for £2 and that she finds she can sell as much broccoli as she can handle at that price. She does not need to reduce her price to increase quantity sold from 200 kilos per week to 300 or 400 or even 500 kilos. The average revenue curve in this case is still the

© ABE and RRC

same as the demand curve, but it reflects this increasing quantity sold at a constant price. This produces the horizontal line graph shown in Figure 3.6.

Marginal Revenue

If a firm is able to maintain a constant price as it increases output, then the additional amount it receives for each extra unit sold is of course that unit's price. In this case the price, which is the same as average revenue, is also the same as the change in total revenue resulting from the sale of the extra unit. The change in total revenue brought about by a small or unit change in the quantity flow of sales is known as the marginal revenue.

Number of TV sets sold per week	**Price per set** £	**Total revenue** £	**Marginal revenue** £
1	600	600	
			550
2	575	1,150	
			500
3	550	1,650	
			450
4	525	2,100	
			400
5	500	2,500	
			350
6	475	2,850	
			300
7	450	3,150	
			250
8	425	3,400	
			200
9	400	3,600	
			150
10	375	3,750	
			100
11	350	3,850	
			50
12	325	3,900	
			0
13	300	3,900	
			–50
14	275	3,850	

Table 3.2: Change in marginal revenue when price is reduced

Marginal revenue is not always the same as the price or average revenue. Remember the example of the leather belts.

There, an increase in sales from 280 to 340 belts per month produced a fall in total revenue. For the change in this output range, the marginal revenue must be negative. The reason is the same as for the fall in total revenue – in order to increase sales, the price had to be brought down. In this case, the revenue gained on the additional quantity sold was not enough to make good the revenue lost for customers who would have been prepared to buy at the higher price.

© ABE and RRC

A simple example will show how marginal revenue can change when price has to be reduced in order to increase the quantity sold. Look at Table 3.2. There are some important features to note about this table. The marginal revenue column has its figures placed midway between the rows. This emphasises that the marginal revenue relates to the change from one output level to the next. On a graph, the marginal revenue is also plotted midway between the output levels. This is shown in Figure 3.7.

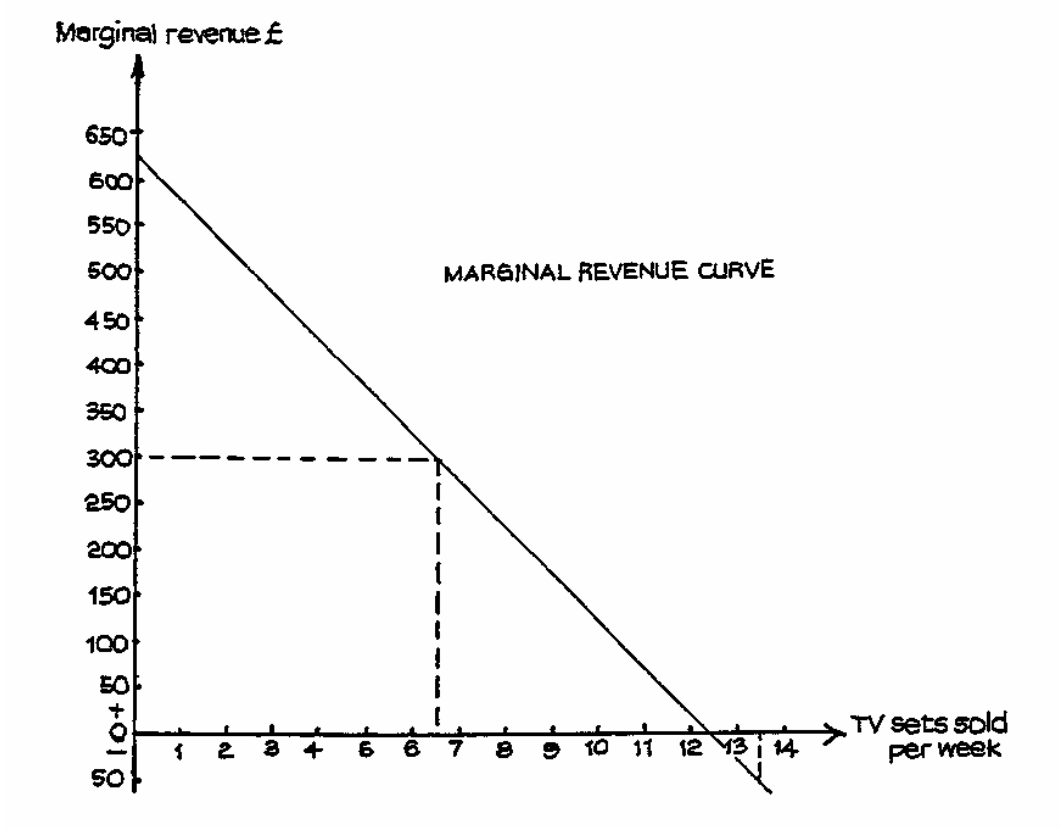


Figure 3.7: Change in marginal revenue when price is reduced

Look carefully at the price and marginal revenue columns. Notice that, as each additional TV set is sold, the price (average revenue) falls £25. The fall in marginal revenue for each additional set is exactly double this – £50.

In Figure 3.8, we see the marginal and the average revenue curves together. Notice that, at each price level, the marginal revenue is exactly halfway between the price axis and the average revenue. Although Figure 3.8 does not continue the average curve until it meets the quantity axis, we can deduce where it would meet if continued in the same straight line. It would meet the quantity axis at 25 TV sets – twice the marginal revenue quantity when marginal revenue equals zero, thus indicating that this supplier would be able to dispose of only 25 sets, even if he did not charge any price at all (i.e. give them away).

The average revenue curve cannot of course pass below the quantity axis, as we do not expect suppliers to pay customers to take their goods. However the marginal revenue curve can pass into the negative area of the graph, and so indicate quantities where continued price reductions would result in an actual fall in total revenue. We can see this clearly from Table 3.2. Marginal revenue remains positive until 12 sets are sold. The increase from 12 to 13 sets does not change total revenue at all, so marginal revenue here is zero. If we continue to reduce price and sell 14 sets, then total revenue falls to £3,850 and marginal revenue indicates the loss as –£50.

© ABE and RRC

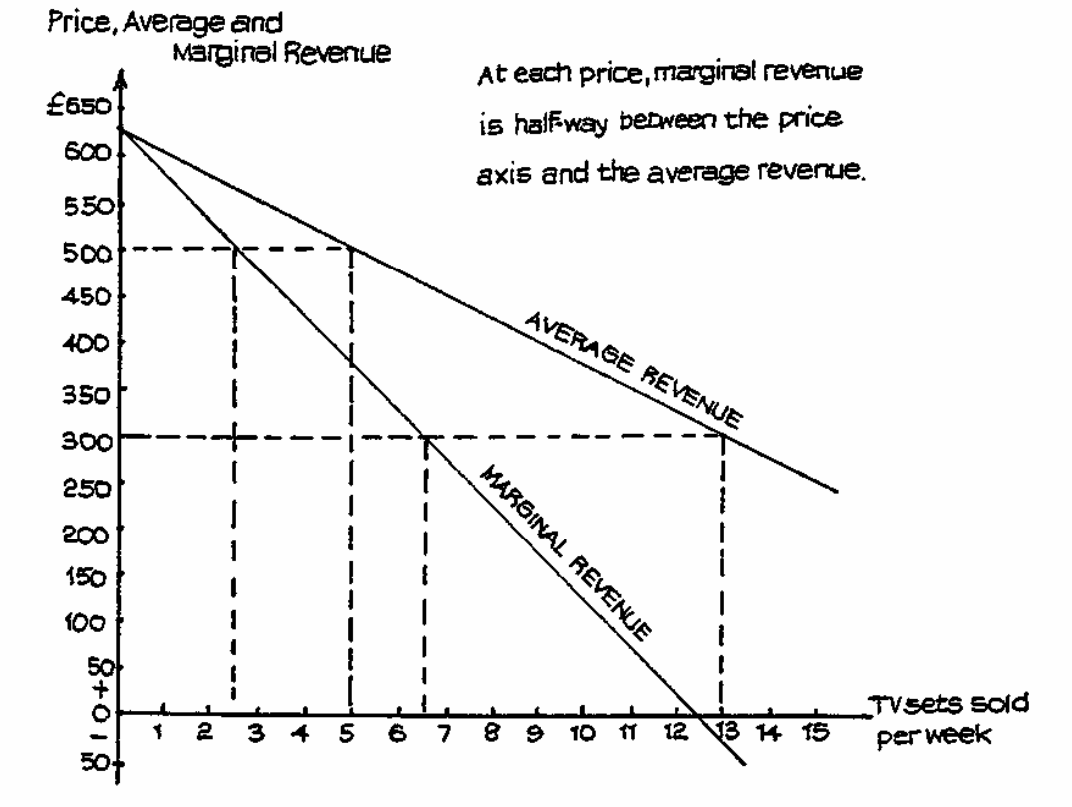


Figure 3.8: Marginal and the average revenue curves

The total revenue curve for this table is shown in Figure 3.9. Compare this with Figure 3.8 and see how the marginal revenue relates to the total revenue at the various numbers of TV sets sold.

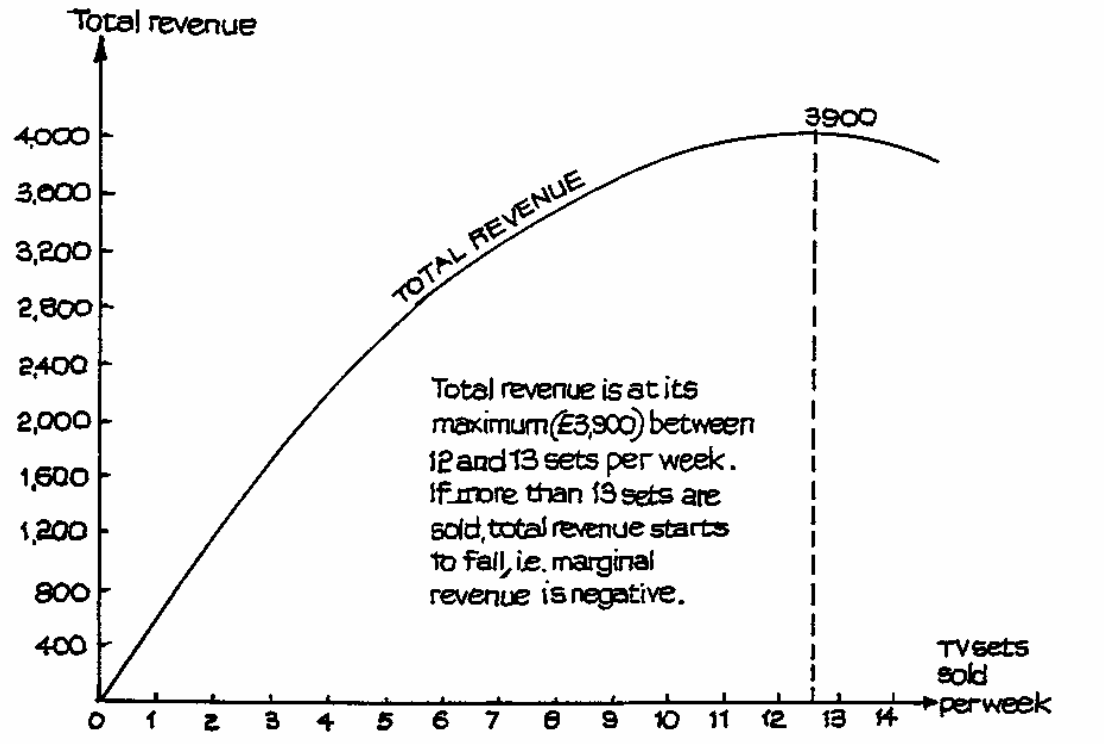


Figure 3.9: Total revenue curve

This example has illustrated an important rule. Whenever we have a linear average revenue curve (i.e. where there is a constant relationship between price and quantity changes resulting in a straight-line graph) then the marginal revenue curve is also linear (a straight line) and always bisects (cuts into two equal halves) the horizontal distance between the price/revenue axis and the average revenue curve.

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. What is the difference between a movement along a demand curve and a shift in the demand curve?
2. Other things remaining unchanged, will an increase in income shift the demand curve for a normal good to the:
   (a) left or
   (b) right?
3. If the cross-price elasticity of demand between two goods is positive are the two goods:
   (a) substitutes or
   (b) complements?
4. What is marginal revenue and how does it change as a firm reduces its price?
5. Complete this statement:
   The other name for a firm's demand curve is its …….
6. A firm is currently selling its product at a price that lies on the inelastic part of its demand curve. In this situation can the firm increase its sales revenue by:
   (a) increasing its price or
   (b) decreasing its price?
7. If a firm's marginal revenue is negative is it operating on:
   (a) the elastic part of its demand curve or
   (b) the inelastic part of its demand curve?
8. To maximise the revenue from placing a sales tax on a good should a government place the tax on a good for which demand is:
   (a) inelastic or
   (b) elastic?

© ABE and RRC

© ABE and RRC

Study Unit 4

Costs of Production

Contents ***Page***

A. Inputs and Outputs: Total, Average and Marginal Product **50**
Factors of Production and Costs 50
Total Product 50
Marginal Product of Labour 51
Average Product of Labour 54

B. Factor and Input Costs **56**
Fixed Costs 56
Variable Costs 57
Total and Average Costs 58
Marginal Costs 59
Long-run Costs 64

C. Economic Costs **65**

D. Costs and the Growth of Organisations **66**
Returns to Scale 66
Economies of Scale 66
Diseconomies of Scale 67
External Economies 68
The Law of Diminishing Returns, Returns to Scale and Economies of Scale 69

E. Small Firms in the Modern Economy **69**
Economies of Scale 69
Services 71
The Role of Small Firms in the Economy 71

© ABE and RRC

Objectives

The aim of this unit is to: discuss the theory of costs, explaining the differences and relationships between various types of cost and distinguishing between the short and long run; solve numerical problems based on cost information; explain and contrast, in words and diagrams, the concepts of economies of scale and returns to scale.

When you have completed this study unit you will be able to:

- explain with reference to appropriate examples, the difference between fixed and variable factors of production
- identify the formulae for, and explain what is meant by, fixed cost, variable cost, marginal cost, average cost and total cost
- solve numerical and/or diagrammatic problems using cost data
- explain, using an appropriate diagram, the relationship between average and marginal cost
- explain, using appropriate examples, the difference between fixed cost and sunk cost
- explain what is meant by economies and diseconomies of scale and relate these concepts to the long-run and short-run average cost curve
- explain what is meant by increasing, constant and decreasing returns to scale and, using real world examples, how each of the these might arise
- compare and contrast the concepts of returns to scale and economies of scale.

A. INPUTS AND OUTPUTS: TOTAL, AVERAGE AND MARGINAL PRODUCT

Factors of Production and Costs

In Study Unit 1 we examined how the factors of production – land, labour and capital – contributed to total production. We also saw that some factors could be regarded as fixed and others could be regarded as variable. This distinction helped to provide us with the important distinction between the short run, when at least one significant production factor was fixed, and the long run, when all factors could be varied.

Total Product

We begin in this section by repeating part of Study Unit 1 and examining what happens when production increases in the short run, when the production factor capital is fixed and when the factor labour is variable. Once again we can take a simple example of a small business which is able to increase its use of labour. For simplicity we can use the term "worker" as a unit of labour, but as remarked before you may wish to regard a worker as a block of worker-hours which can be varied to meet the needs of the business.

Suppose the effect of adding workers to the business is reflected by Table 4.1, where the quantity of production is measured in units and relates to a specific period of time, say, a month. The amount of capital employed by the business is fixed.

© ABE and RRC

Number of workers	**Quantity of production** *(units per month)*
1	30
2	70
3	120
4	170
5	220
6	260
7	290
8	310
9	320
10	320
11	310

Table 4.1: Number of workers and quantity of production

The quantity of production (measured here in units produced per month) which is shown as a graph in Figure 4.1 is of course the *total product*. In this example total product continues to rise until the tenth worker is added to the business. This worker is unable to increase total product. Given the fixed amount of capital, no further increase in productive output is possible. The addition of an eleventh worker would actually cause a fall in production.

Marginal Product of Labour

Now examine the amount of change to the total product as each additional worker is added to the business. Table 4.2 shows this change in the third column, which is headed *marginal product*. Strictly speaking, this is the *marginal product of labour* because it results from changes in the amount of labour (workers) added to the business.

© ABE and RRC

Number of workers	Quantity of production *(units per month)*	Marginal product of labour *(additional units per month)*
0	0	
		30
1	30	
		40
2	70	
		50
3	120	
		50
4	170	
		50
5	220	
		40
6	260	
		30
7	290	
		20
8	310	
		10
9	320	
		0
10	320	
		–10
11	310	

Table 4.2: Adding marginal product of labour

The marginal product of labour is the *change in total product resulting from a change in the amount of labour employed*. It is called marginal because it is the change at the edge; the term "marginal" is used in economics to denote a change in the total of one variable which results from a single unit change in another variable. Here the total is quantity of production resulting from changes in the number of workers employed.

The marginal product column shows the difference in the total product column at each level of employment. Notice that the marginal value is shown midway between the values for total product and the number of workers. This is because it shows the change that takes place as we move from one level of employment (i.e. adding an additional worker) to the next. In Figure 4.1 the marginal product is represented by the vertical distance between each step in production as each worker is added.

The sum of the marginal product values up to each level of worker is equal to the total product at that level.

© ABE and RRC

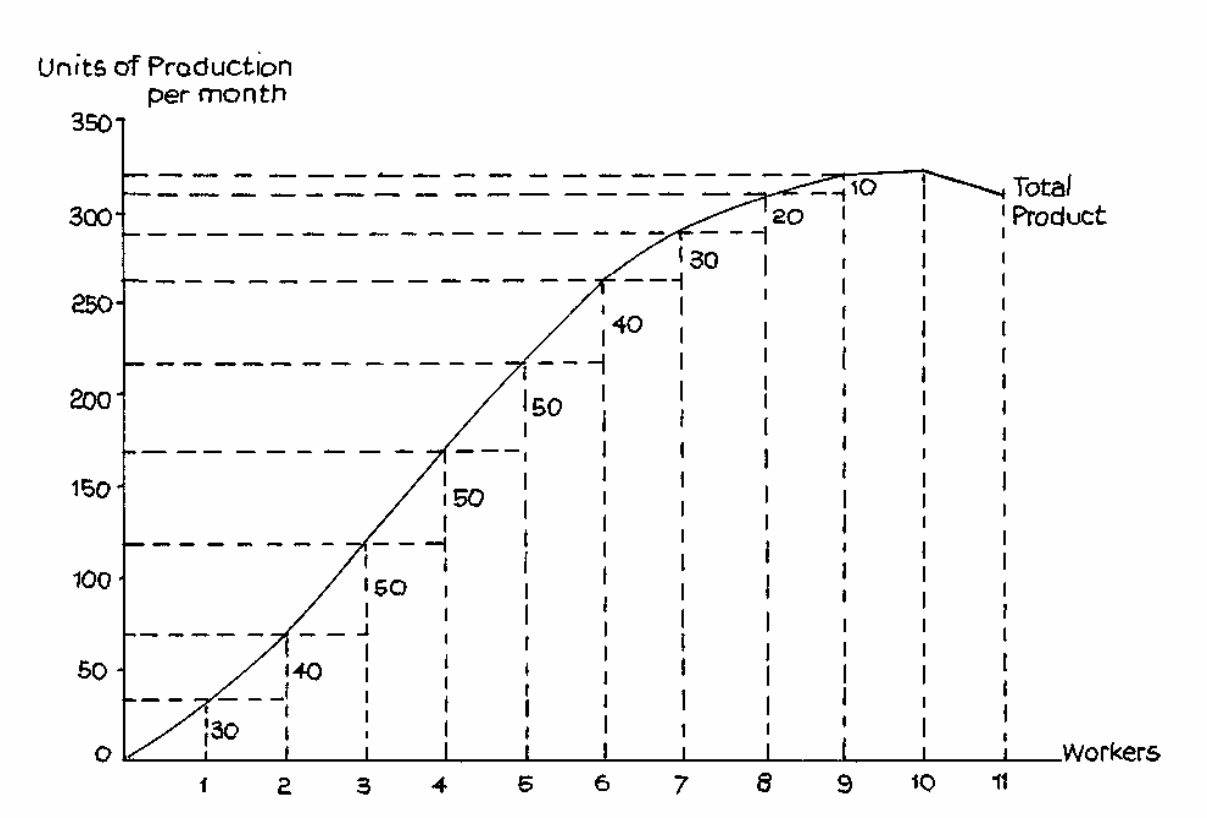


Figure 4.1: Illustrating marginal product

Notice how the value for marginal product changes as total product rises: one worker alone can produce 30 units but another enables the business to increase production by 40 units and one more by 50 units. There are many ways in which this increase might be achieved, e.g. by specialisation and by freeing the manager to improve administration, purchasing and selling. However, these increases cannot continue and the additional third, fourth and fifth workers all add a constant amount to production. Thereafter, further workers, while still increasing production, do so by diminishing amounts until the tenth worker adds nothing to the total. At this level of labour employment production has reached its maximum, and the eleventh worker actually provides a negative return – total production falls. Perhaps people get in each other's way or cause distraction and confusion. If the business owner wishes to continue to expand production, thought must be given to increasing capital through more buildings and/or equipment. Short-run expansion at this level of capital has to cease. Only by increasing the fixed factors can further growth be achieved.

As remarked in Study Unit 1, this particular example is purely fictional – it is not based on an actual firm: but neither is the pattern of change in marginal product accidental. The figures are chosen deliberately to illustrate some of the most important principles of economics, the so-called laws of *varying proportions* and *diminishing returns*. It has been constantly observed in all kinds of business activities that when further increments of one variable production factor are added to a fixed quantity of another factor, the additional production achieved is likely first to increase, then to remain roughly constant and eventually to diminish. It is this third stage that is usually of the greatest importance, this is the stage of *diminishing marginal product*, more commonly known as *diminishing returns*. Most firms are likely to operate under these conditions. It is during this stage that the most difficult managerial decisions, relating to additional production and the expansion of fixed production factors, have to be taken.

Of course it must not be assumed that firms will seek to employ people up to the stage of maximum product when the marginal product of labour equals zero, or on the other hand that they will not take on any extra employees if diminishing returns are being experienced.

© ABE and RRC

The production level at which further employment ceases to be profitable depends on several other considerations, including the *value of the marginal product*. This depends on the revenue gained from product sales, and the cost of employing labour, which is made up of wages, labour taxes and compulsory welfare benefits. The higher the cost of employing labour, the less labour will be employed in the short run and the sooner will employers seek to replace labour by capital in the form of labour-saving equipment.

Average Product of Labour

The average product of labour employed is found simply by dividing the total product at any given level of employment by the number of workers (or some unit of worker-hours). For reasons which by now should be starting to become apparent to you, the average product of labour, though a measure easily understood and used by many business managers and their accountants, is less important than the marginal product. However, Table 4.3 adds average product to our earlier statistics, and Figure 4.2 shows both marginal and average product in graphical form.

Number of workers	**Quantity of production** *(units per month)*	**Marginal product of labour** *(units per month)*	**Average product of labour** *(units per month)*
0	0		
		30	
1	30		30.00
		40	
2	70		35.00
		50	
3	120		40.00
		50	
4	170		42.50
		50	
5	220		44.00
		40	
6	260		43.33
		30	
7	290		41.43
		20	
8	310		38.75
		10	
9	320		35.56
		0	
10	320		32.00
		–10	
11	310		28.18

Table 4.3: Adding average product

© ABE and RRC

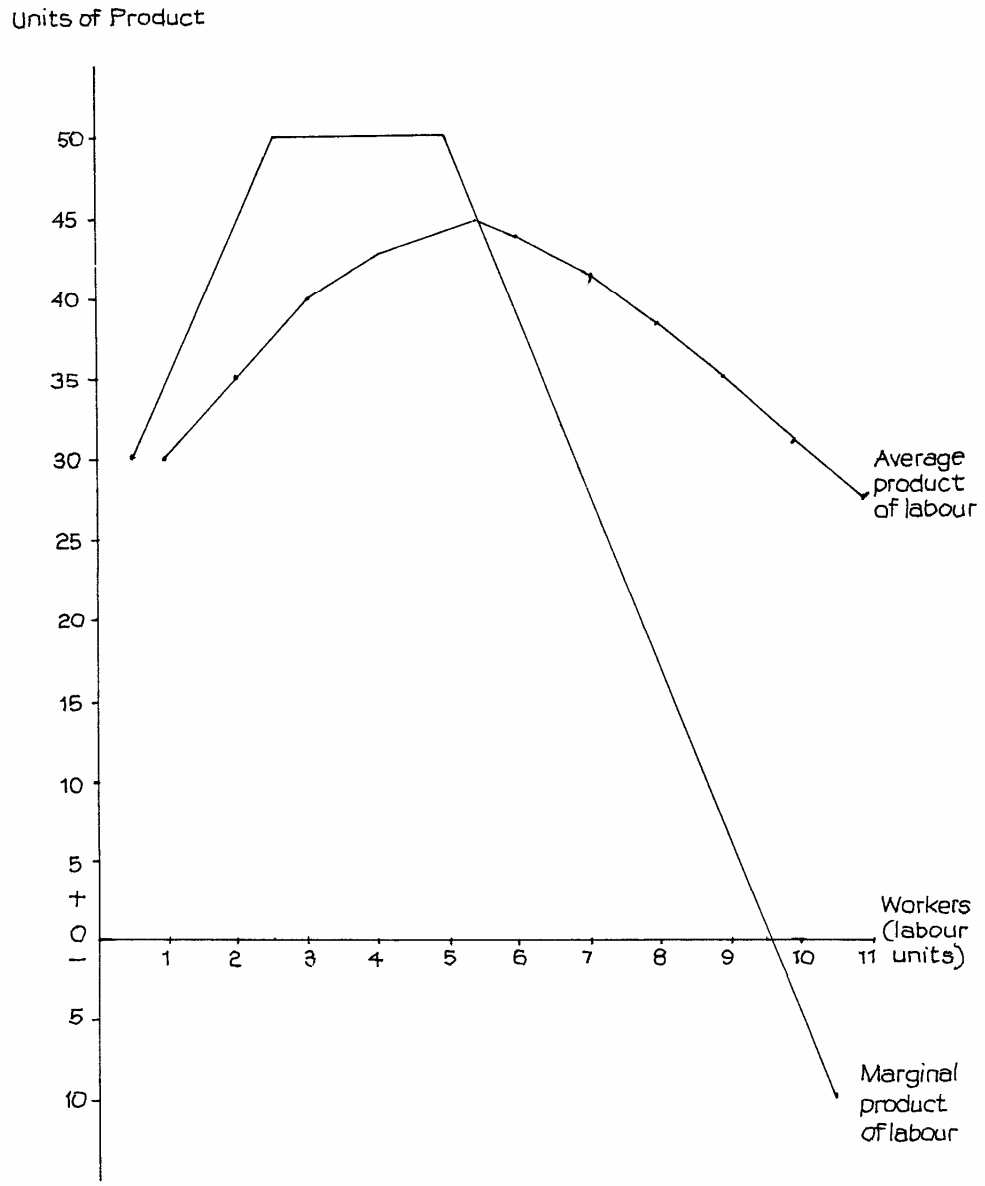


Figure 4.2: Marginal product and average product curves

The falling marginal product curve intersects the average product curve at about the 5th worker. Average product then starts to fall because for more workers marginal product is below average product.

Notice the relationship between average and marginal product. Average product continues to rise until it is the same as the falling marginal product, then it falls. This must happen as can easily be proved mathematically, and you can see it for yourself if you take any set of figures where marginal product continues to diminish.

© ABE and RRC

B. FACTOR AND INPUT COSTS

The payments made to the owners of production factors in return for their use in the process of production are of course the costs of production, which the production organisation (firm) has to pay in order to produce goods and services. More strictly these are termed the *private* production costs. These factor payments, in very general terms, are rent to the owners of land, interest to the owners of capital and wages to the providers of labour. Disregarding land for the sake of using very simple models, we can, initially, regard capital as the major fixed production factor and labour as the variable factor.

Fixed Costs

Fixed costs are the costs of the fixed factors, i.e. those elements which are not being increased as production or output is being raised. The total fixed costs for a given range of output can be illustrated in the simple graph shown in Figure 4.3.

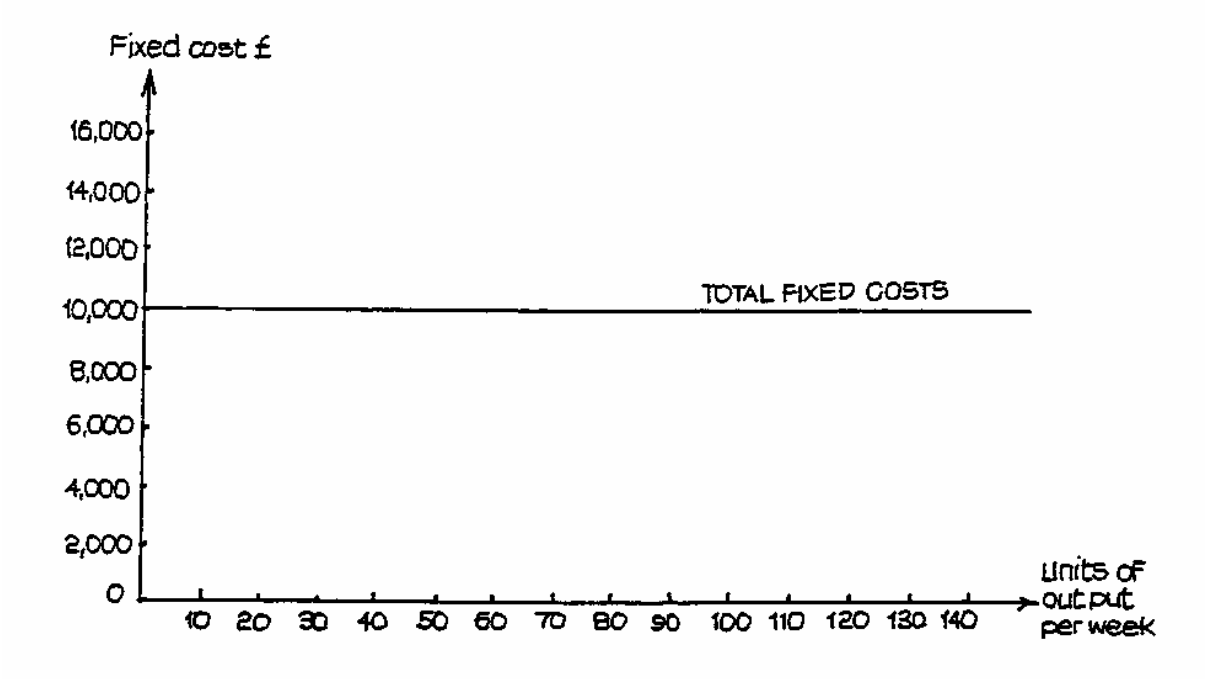


Figure 4.3: Total fixed costs

Examples of fixed costs include rent for land or buildings, rental charge for telephone or telex, business rates, salary of a manager, and fees for a licence to make use of another company's patent. All these costs can change, but the point is they do not change as production level changes. The cost has to be met, whatever the level of output and sales.

The graph of average fixed costs, i.e. total fixed costs divided by the number of units of output produced, is shown in Figure 4.4. This is based on the fixed costs of £10,000 assumed in Figure 4.3. Notice the steep fall at the lower levels of output, and the much more gentle slope of the curve at higher levels. Between 140 and 150 units of output per week, the fall in average fixed costs is only from approximately £71 to £67.

© ABE and RRC

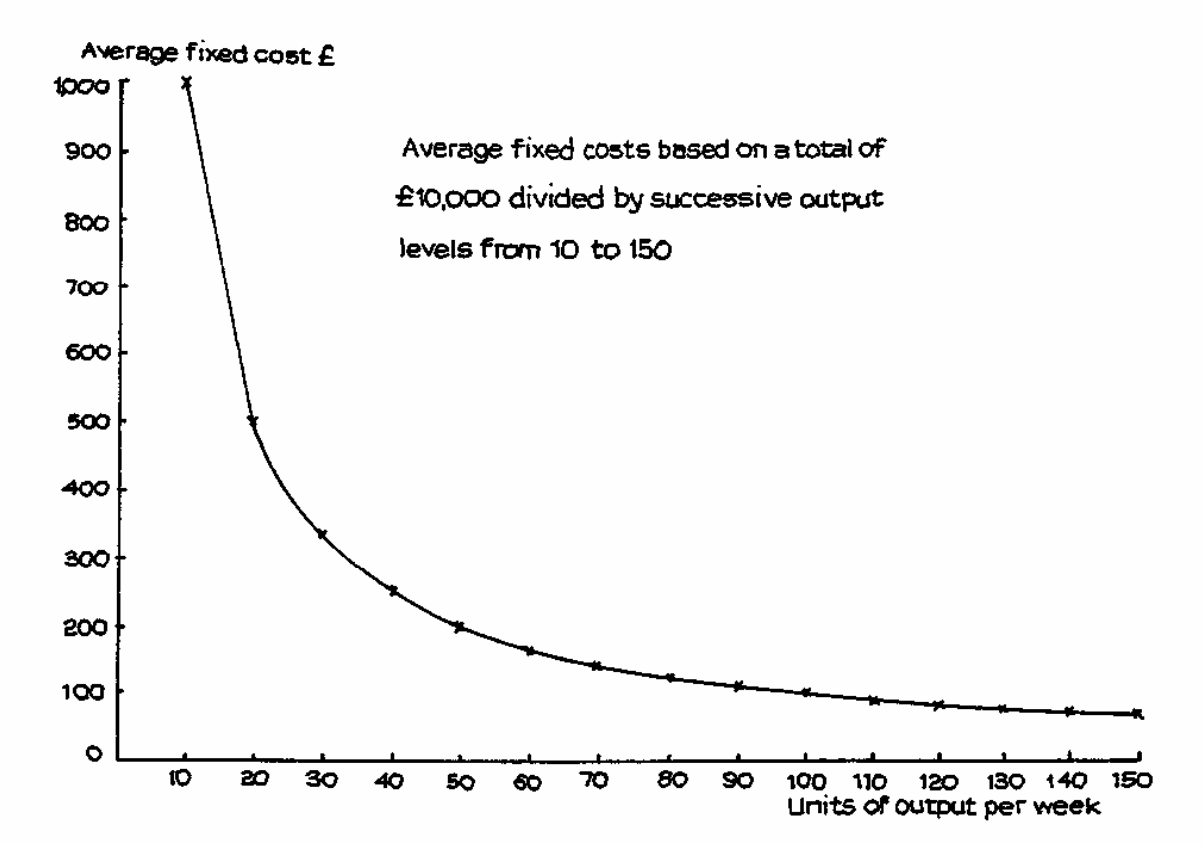


Figure 4.4: Average fixed costs

Variable Costs

The behaviour of variable costs depends on the pattern of production returns. If production is rising faster than the input of variable elements, then costs are increasing less than proportionally to the rise in output. This is because each extra unit of input is adding more to production than it is to cost. This is possible at the lower levels of production represented by the section of graph 0a in Figure 4.5.

Later, costs are likely to rise in the same proportion as output – this being the stage of constant returns, shown between output levels 0a and 0b. Then, as we reach the level of diminishing returns, costs rise faster (more steeply) than production. This is shown beyond level 0b.

© ABE and RRC

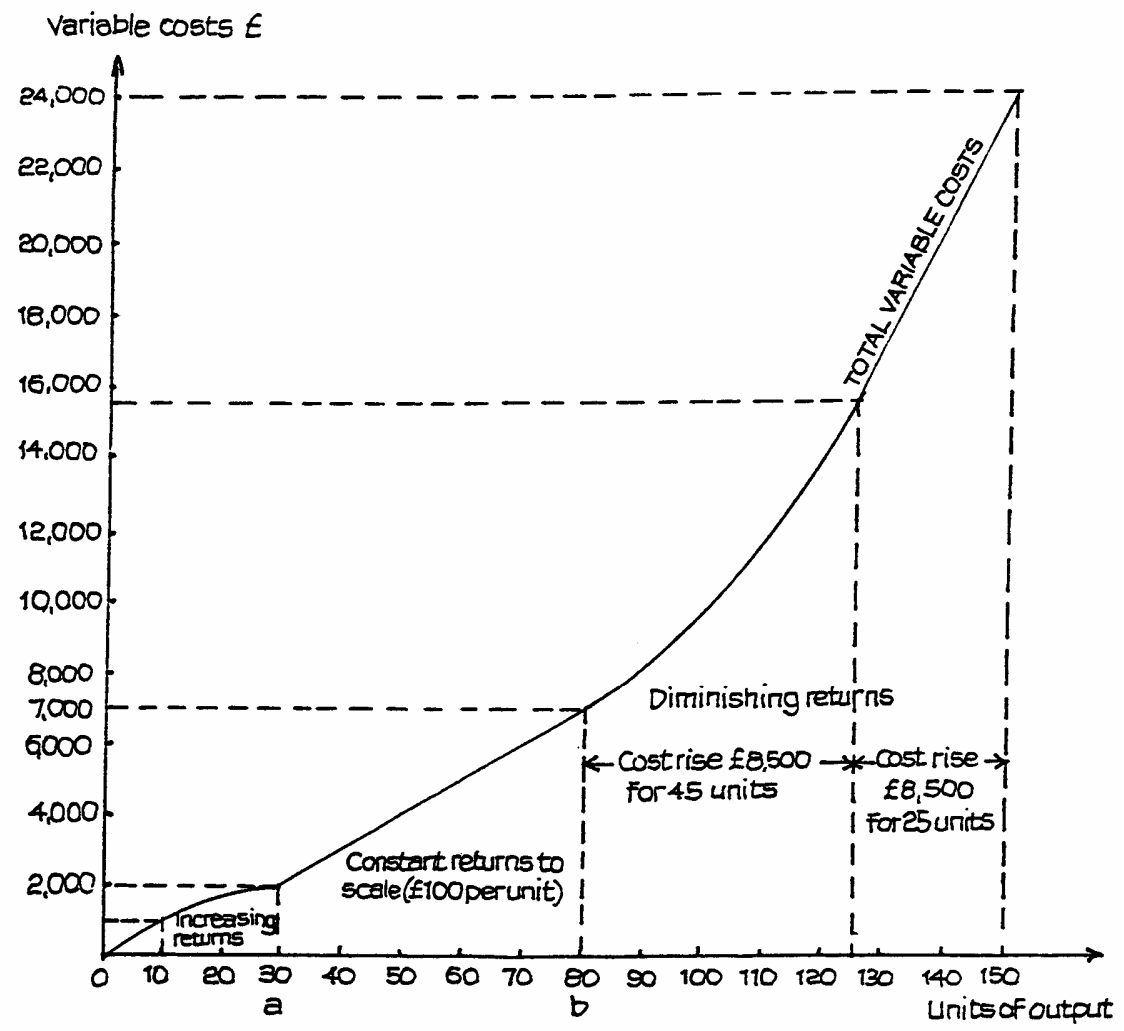


Figure 4.5: Total variable costs

Total and Average Costs

If we combine fixed and variable costs, we obtain total costs. So, if we combine Figure 4.3 (which shows total fixed costs) with Figure 4.5, we obtain the graph of total costs. This is shown in Figure 4.7.

From the total costs we can obtain average total costs, simply by dividing the total by each successive level of output. Average total costs are often referred to just as average costs. Figure 4.6 shows the graph of average total costs, which has been derived from the total cost curve shown in Figure 4.7.

Notice how the shape of the average cost curve at the lower levels of output is very similar to that of the average fixed cost curve in Figure 4.4. This is because, at these levels, fixed costs form a high proportion of total costs. As fixed costs become a smaller proportion of total costs, the curve falls much less steeply. In this illustration, it reaches its lowest point a little below the 110 units per week output level and then begins to rise, as variable costs become steeper in response to diminishing returns to scale.

This is the typical shape of the curve in the short run (that is, while "fixed" costs remain the unchanged). Because it falls to a minimum point and then rises, it is often referred to as the "U-shaped" average cost curve, although as you can see, a more accurate description is that of an L with its toe turned upwards. Only if there are particularly severe increasing costs

(diminishing returns) to scale, and fixed costs are a very small proportion of total costs, will the second half of the "U" be at all steep; the efficient firm should never allow itself to reach this position.

The modern firm is more likely to have a high proportion of fixed to total costs, because of the swing from labour to labour-saving machinery. This movement is described as production becoming more and more capital-intensive. In this case, we can expect the average total cost curve increasingly to resemble the average fixed cost curve.

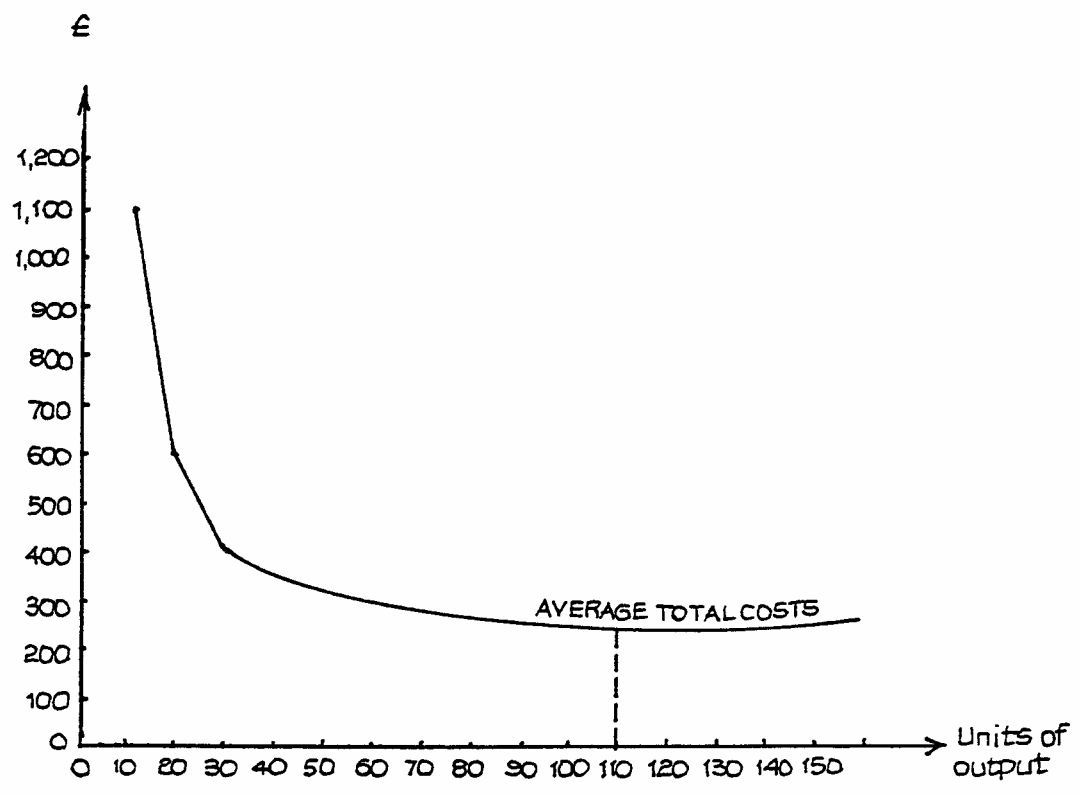


Figure 4.6: Average total costs

Marginal Costs

You have already met marginal product, marginal utility and marginal revenue – the change in total output, utility or revenue as output changes. You will not then be surprised to know that marginal cost is the change in total cost as output changes. Once again, we relate this change to a single unit of output so that, if we are moving in steps of ten (as in our cost example so far), we shall have to divide any change from one forward step to the next by ten.

© ABE and RRC

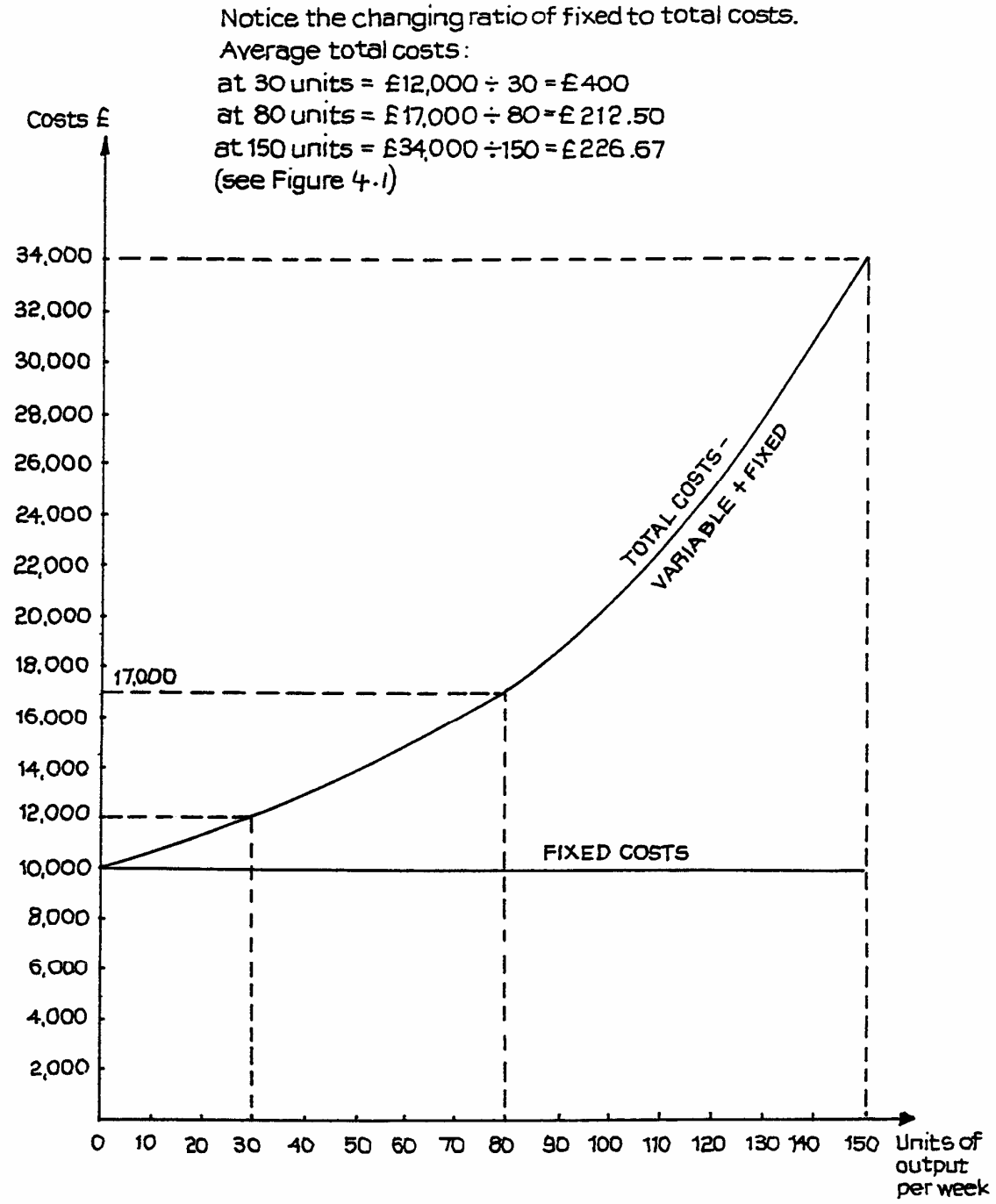


Figure 4.7: Total cost curve

Table 4.4 is a table of total (fixed plus variable) costs which correspond to our previous graphs. In this table, further columns have been added to show the change in total cost between each output step of ten units per week, and then division by ten to produce the marginal cost. Notice that the figures of the marginal cost column have been placed midway between the figures of the other columns, to emphasise that they relate to a change from one output level to the next.

© ABE and RRC

(1)	(2)	(3)	(4)
Quantity	**Total cost**	**Changes in total cost from one quantity level to the next**	**Marginal cost** (column 3 divided by 10)
(units per week)	*£*	*£*	*£*
0	10,000		
			100
10	11,000	1,000	
			60
20	11,600	600	
			40
30	12,000	400	
			100
40	13,000	1,000	
			100
50	14,000	1,000	
			100
60	15,000	1,000	
			100
70	16,000	1,000	
			100
80	17,000	1,000	
			115
90	18,150	1,150	
			135
100	19,500	1,350	
			165
110	21,150	1,650	
			210
120	23,250	2,100	
			275
130	26,000	2,750	
			355
140	29,550	3,550	
			445
150	34,000	4,450	

Table 4.4: Cost table

On a graph, the marginal cost is plotted at the midpoints of the various output levels. You will see that this has been done in Figure 4.8, which illustrates the marginal costs shown in Table 4.4.

© ABE and RRC

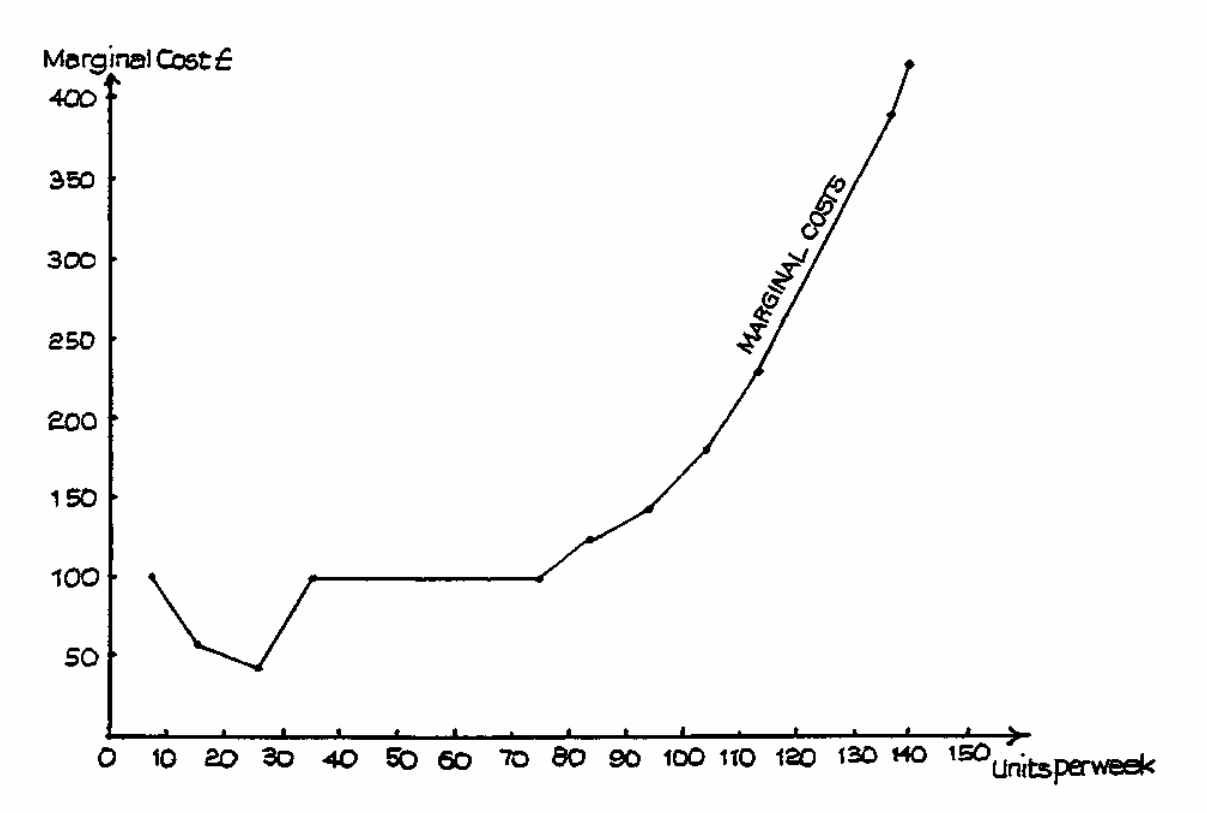


Figure 4.8: Marginal costs

In Figure 4.9, the marginal cost graph has been combined with the average cost graph. Notice where these two curves intersect.

The rising marginal cost curve cuts the average cost curve at roughly 110 units per week. This is the output level which we have already noted as the lowest level of the average total cost curve. This illustrates a rule that you must remember: *the rising marginal cost curve always cuts the average cost curve at its lowest point.* If you think a little, you will see that it must do that. If the cost of the last unit to be produced is less than the average up to that point, then the new average will be a little lower. If the cost of the last unit is higher than the average up to that point, then the new average will be a little higher.

Experiment with some simple figures and you will see that this must always be true. Remember this relationship, and always show the correct intersection when you draw graphical illustrations.

© ABE and RRC

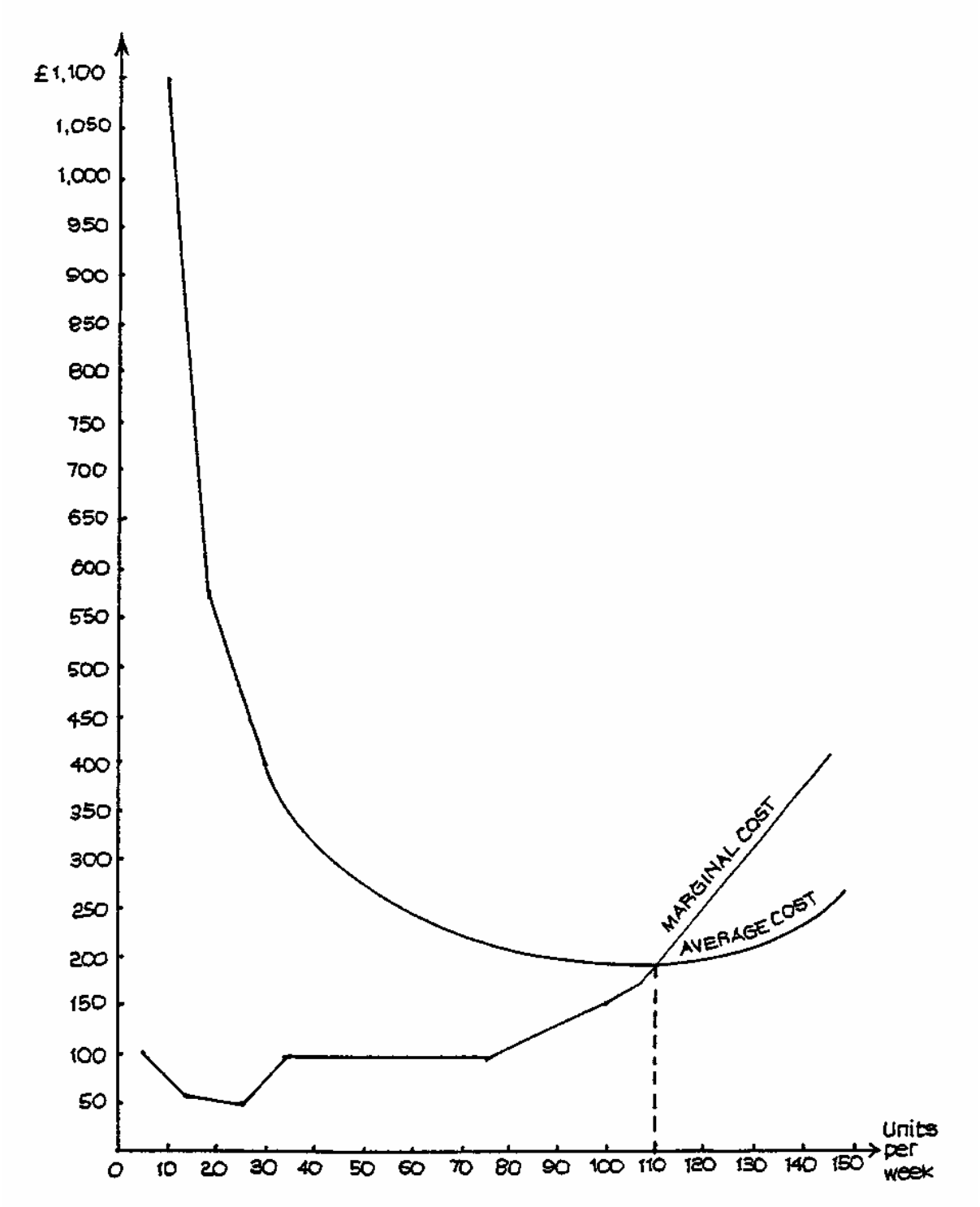


Figure 4.9: Marginal cost and average cost

© ABE and RRC

Long-run Costs

In the long run all factors of production may be increased: no costs are completely fixed. In practice of course the factors which are fixed in the short run will be increased in definite stages, perhaps when a new factory is built or when new technology introduced, etc. The graph of fixed costs in the long run, therefore, appears as in Figure 4.10.

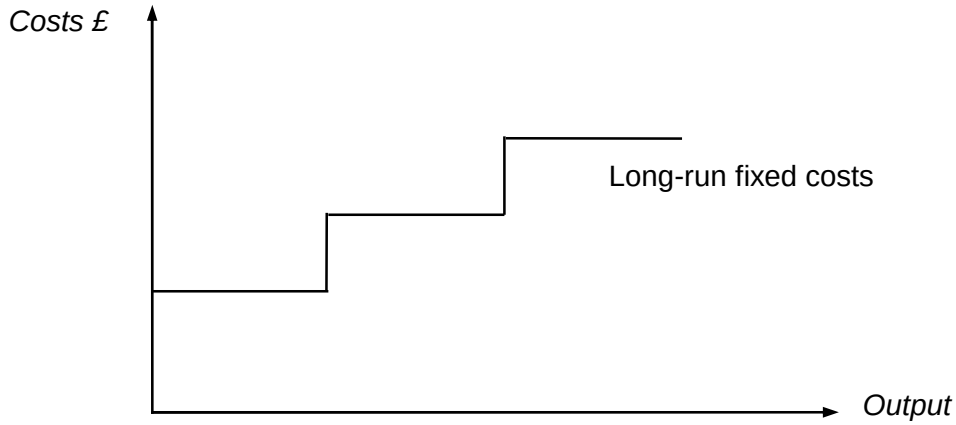


Figure 4.10: Fixed costs in the long run

The effect of this on the average total cost curve in the long run is shown in Figure 4.11.

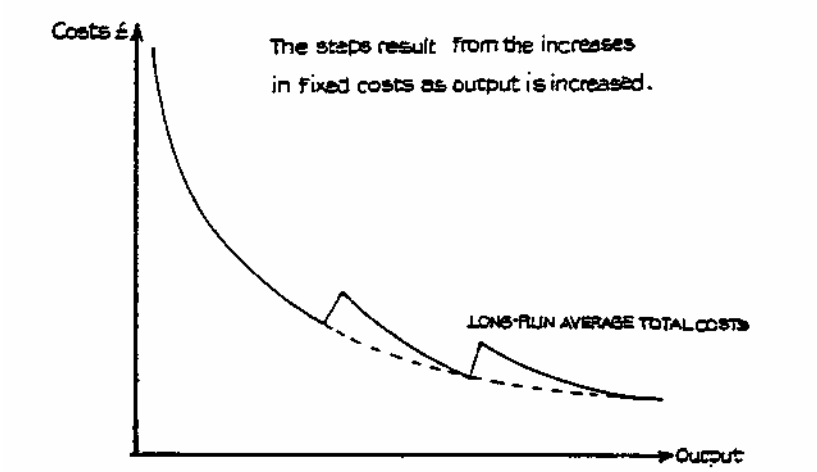


Figure 4.11: Effect of long-run fixed costs on total cost

The "flat" part of the average cost curve is prolonged. The question is whether this merely stretches the average cost curve – delaying the point of eventual diminishing returns and the rise of the U shape – or whether it can be continued indefinitely, in order to prevent the U shape completely and make the long-run average cost curve L-shaped.

The relationship between short-run and long-run average cost curves is sometimes shown as in Figure 4.12. This emphasises the fact that one reason for the increase in fixed factors and costs is to overcome the effect of short-run diminishing returns.

© ABE and RRC

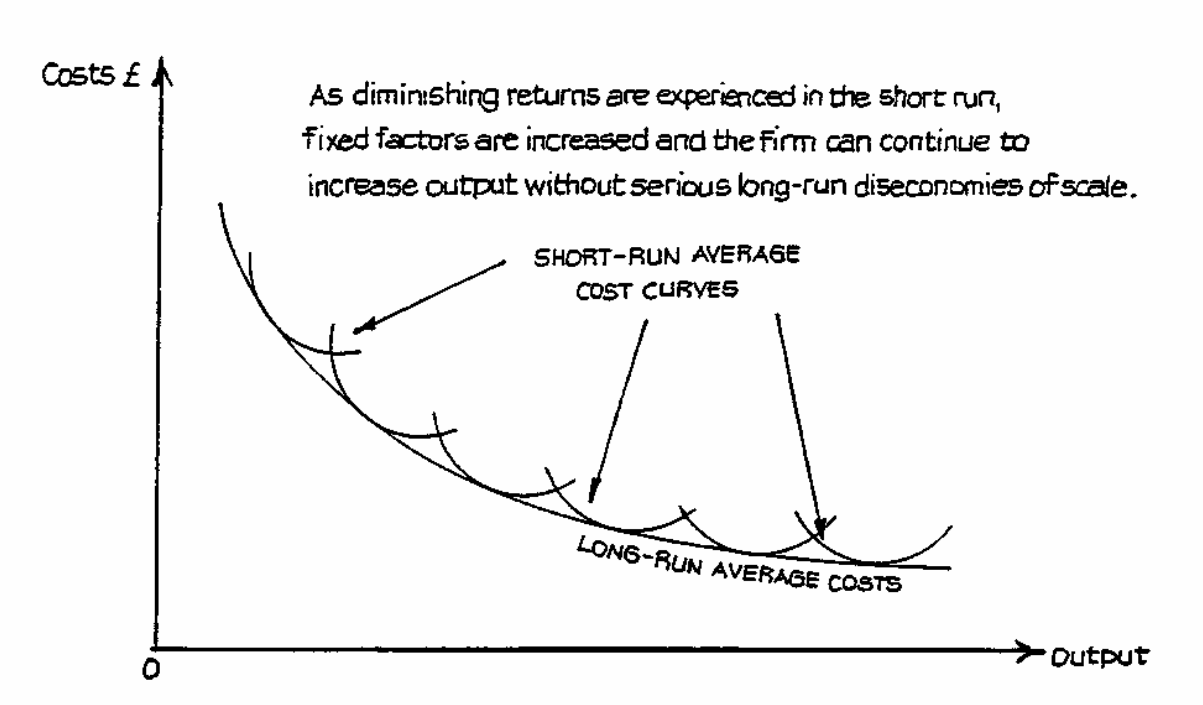


Figure 4.12: Relationship between short-run and long-run average cost curves

C. ECONOMIC COSTS

We are now beginning to see production costs from a variety of angles.

- **Opportunity Costs**

  These were identified in Study Unit 1. They may be defined as the cost of using resources in one activity measured in terms of the lost opportunity of using them to produce the best alternative that had to be sacrificed.

- **Absolute Costs**

  These are the full costs of the factors used in the activity under consideration. They may be measured in monetary terms but the real absolute cost is best measured by the actual quantity of factors used, e.g. the amount of land or the numbers of people employed.

- **Private Costs**

  These are the costs actually paid by the producer to the owners or providers of the production factors employed. They are the costs usually taken into account by the accountant and are measured in monetary terms, since the accountant has to account for the use of whatever finance has been entrusted to the production organisation. We have been looking at these costs in this study unit and have also examined the important distinction between fixed and variable costs.

- **External Costs or Social Costs**

  These are the indirect costs imposed on other firms or individuals as a consequence of the process of production by firms. Because these costs are imposed on others in society they are also known as social costs to distinguish them from private costs. Producers have to pay for the direct costs they incur in production (their private costs), but do not take account of the external costs they may also be imposing on society. The main source of such external costs is pollution of the environment. If an electricity supply company burns coal or gas to generate electricity the company pays the market price for the coal or gas it burns as its main input into the production of electricity (its

© ABE and RRC

main variable factor of production). Unfortunately for society, and the world environment, the large-scale burning of coal or gas not only generates electricity, it also releases large amounts of pollution into the atmosphere, especially carbon which is a major factor in global warming. Unless governments take action to deal with this problem, by imposing taxes on the use of combustible fuels to generate electricity, the social cost is not taken into account by electricity producers when they decide which fuel and how much of it to use.

We will look at these issues in more detail in Unit 6.

D. COSTS AND THE GROWTH OF ORGANISATIONS

Returns to Scale

We have already seen the results of increasing inputs of a variable factor when at least one other production factor is held constant. We saw that this was likely to bring about first increasing, followed by constant and then diminishing marginal returns. However we have also pointed out that, in the long run, all factors can be increased: there is the possibility of economies of scale resulting for the continued growth in size of the firm. We must now look at this possibility more closely, but first we must be clear as to the meaning of returns to scale when all factors are being increased. If a given proportional increase in factors results in a larger proportional increase in output, then the firm is enjoying increasing returns, or economies of scale. For example this would be the case if a 10 per cent increase in factor inputs produced a 20 per cent increase in production output.

If the proportional increase in output is the same as the proportional increase in factor inputs (e.g. when a 15 per cent increase in factors produces a 15 per cent increase in output) then the firm is experiencing constant returns. However if a 15 per cent increase in factor inputs produces less than a 15 per cent increase in output (only 10 per cent, say) then the firm is suffering decreasing returns, or diseconomies of scale.

Economies of Scale

Real scale economies, as defined here, should be distinguished from purely pecuniary or monetary economies. The latter do not represent a more efficient use of factors; rather they are the result of the superior bargaining power of the large firm in the market. For instance, a large customer can often gain discounts greater than can be justified on the grounds of savings in delivery or distribution costs. Or workers in some large firms may be willing to accept a lower wage in return for what is believed to be greater security of employment or the social prestige of working for a famous organisation. Real economies – the genuine efficiencies in the use of production factors resulting from growth in the scale of activities – can be identified in the following main areas.

- **Labour Economies**

  Labour economies result from greater opportunities for the division of labour which increase with the skills of the workforce, save time and allow greater mechanisation. The automated assembly line in modern motor vehicle assembly is an extreme example of this.

- **Technical Economies**

  Technical economies result chiefly from the use of specialised capital equipment. Large firms are able to make use of equipment that could not be fully employed by smaller operations, and large firms are also able to support reserve machines to avoid disruption following breakdown. A small firm, using three machines, adds one-third to its capital cost if it tries to add a further machine to keep in reserve. A large firm employing 20 machines adds only one-twentieth if it decides to do likewise.

© ABE and RRC

- **Marketing Economies**

  Very great economies are available from large-scale advertising. A television commercial using top stars is very expensive to make, but the cost per potential customer is very low if essentially the same film can be shown in several different countries. Large firms can also afford to keep very skilled marketing specialists fully employed.

- **Financial Economies**

  Large firms are able to obtain finance from markets that are denied to small firms, and multinationals can raise money in many different countries. Nevertheless, although financial economies still exist, we do have to recognise that finance markets have, in recent years, become more responsive to the needs of smaller enterprises.

- **Distribution and Transport Economies**

  Transport movements and the location of depots can be carefully planned by large organisations, so that vehicles and storage space are used efficiently.

- **Managerial Economies**

  Managerial economies arise from the employment of specialised managers and managerial techniques. However many of these techniques have been developed in order to overcome the problems of managing large organisations, and many economists suggest that managerial economies of scale are often exaggerated and difficult to achieve in practice.

Diseconomies of Scale

Diseconomies of scale are usually associated with the problems arising out of the management and control of large organisations. Formal communication systems are necessary but are expensive to maintain. Whereas the manager of a small organisation can see what is going on around him or her in the course of daily work, the manager of a large firm may have to establish an inspection system to obtain equivalent information – which is unlikely to be as reliable.

There can also be a loss of control over managers at the lower levels of the managerial pyramid. These managers may then pursue their own private objectives (e.g. building up the power of their own department) at the expense of efficiency and profitability.

So diseconomies of scale are mostly managerial. If diseconomies just balance economies, e.g. when a 10 per cent increase in factor inputs produces the same 10 per cent increase in production output, the long-run average cost curve will have the L shape of Figure 4.13. If economies of scale continue roughly to balance diseconomies, this shape may be retained over a long period. However if diseconomies start to rise substantially, then the long-run average cost will again start to rise.

© ABE and RRC

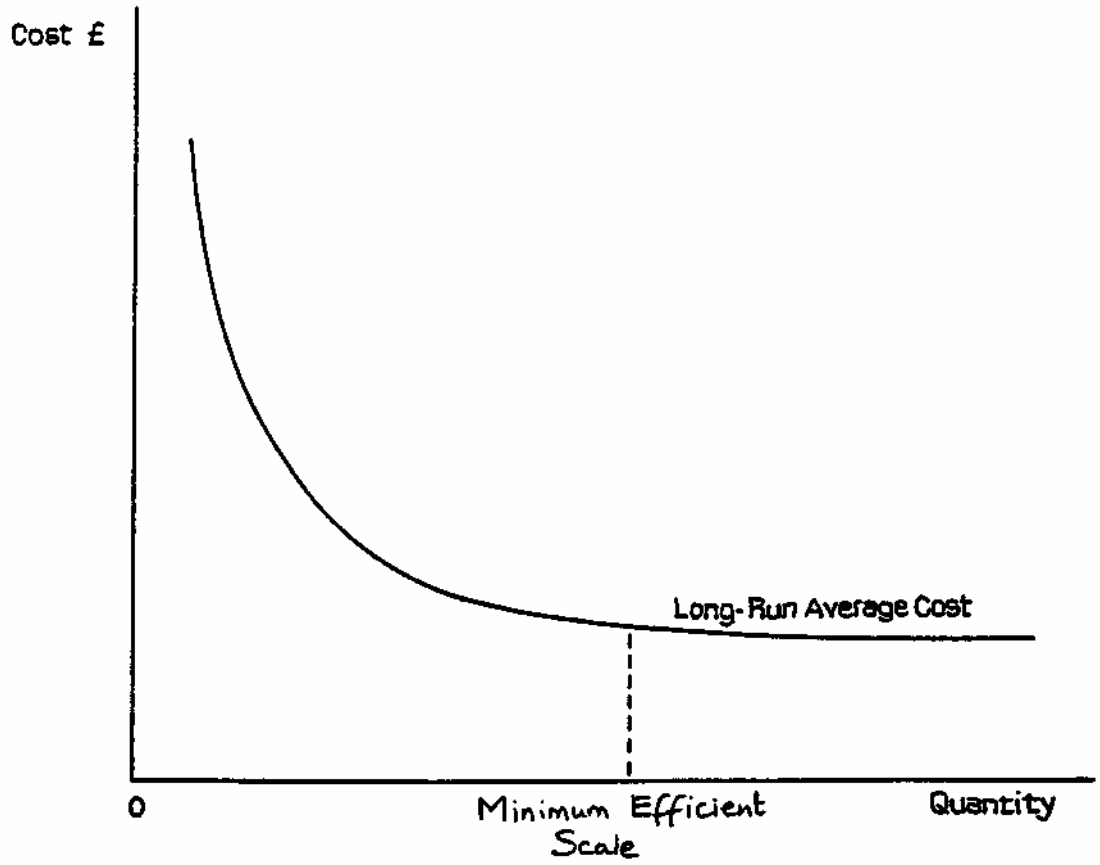


Figure 4.13: Long-run average cost curve

Notice here the position of what is called the *minimum efficient size (or scale) (MES)*, also known as the minimum optimum scale (MOS). Up to this output level there are significant gains from internal economies of scale. Firms operating below the MES are at a cost disadvantage when competing against those operating up to or beyond that size. However beyond the MES further cost savings are not significant, and there is no cost advantage in further growth. On the other hand the shape of the curve can change as firms learn how to overcome sources of inefficiency, in particular managerial inefficiency, especially when new managerial skills and communication technology are introduced. It is possible to control very large firms today in ways that would have been impossible half a century ago. Jet travel and modern telecommunications, not to mention computers and microelectronics, have transformed management techniques.

External Economies

The economies of scale listed earlier all apply to the individual firm; they are known as internal economies of scale. There are other economies that are external to the firm. These arise when an industry grows large or when business firms congregate in a particular area. External economies usually arise from the development of specialised services available to many firms. For example, an area containing numbers of small engineering companies may provide opportunities to support one or more specialised toolmakers. Each engineering company can call on the specialist, without having to carry the full cost of having its own specialised department. External economies help small firms to survive in competition with larger organisations. However, if one or two companies become dominant and they internalise these economies by setting up their own specialised departments which they are large enough to keep fully employed, then the external economies may be lost to the smaller firms, which can then no longer survive in the market.

© ABE and RRC

The Law of Diminishing Returns, Returns to Scale and Economies of Scale

The shape of a firm's average cost curve in the short run is determined by its fixed factors of production, usually machinery, buildings or land, and the unavoidable operation of the law of diminishing returns. At some point as a firm tries to squeeze out yet more output from its fixed physical capacity by application of additional workers and materials, it will start to experience diminishing marginal product and its unit cost of production will start to rise at an increasing rate. The firm's short-run average cost curve will thus always turn up at some point and have a U shape. The downward sloping portion of the U-shaped cost curve is not due to economies of scale, because the scale or size of the firm is fixed in the short run. Likewise, the upward sloping portion of the U-shaped cost curve is not due to diseconomies of scale, because the scale or size of the firm is fixed in the short run. The shape is determined by what happens to the marginal product of successive inputs of variable factors to a fixed factor – the *law of diminishing returns*.

Economies and diseconomies of scale relate to what happens to a firm's average or unit cost of production as the firm increases its output by expanding the availability of all the factors of production it needs. That is, economies and diseconomies determine the shape of a firm's long-run average cost curve. If a firm benefits from economies of scale, as it expands in the long run it experiences increasing returns to scale as its average cost of production falls. In contrast, if a firm suffers from diseconomies of scale, as it expands it will experience decreasing returns to scale as its average cost of production increases. These relationships are summarised in Table 4.5.

Neither economy nor diseconomy of scale	=	Constant returns to scale	=	Constant unit cost
Economy of scale	=	Increasing returns to scale	=	Decreasing unit cost
Diseconomy of scale	=	Decreasing returns to scale	=	Increasing unit cost

Table 4.5: Relationships between economies of scale, returns to scale and unit costs

E. SMALL FIRMS IN THE MODERN ECONOMY

It is sometimes assumed that because of economies of scale, large firms are always likely to be more efficient and produce at lower cost than small firms. If this were true, small firms would be much less numerous than they are. Of course, one reason for their survival is that the definition of a small firm tends to change in time. As the average size of the firm grows, so firms which would have been considered large become classified as small. Moreover, if we take as the main qualification to be considered a small firm, the requirement that the whole enterprise is controlled by a small group of employer-managers, continued advances in technology, including information technology, enable one or two people to control larger enterprises. This means many more firms can now grow larger but remain, in fundamental respects, small.

Economies of Scale

A closer look at economies of scale shows that large firms are not always inevitable. If we assume that the typical successful large company has an L-shaped cost curve, this can still cover a number of different possibilities.

© ABE and RRC

Figure 4.14 shows two possible long-run average cost curves. It shows that each reaches a point where further cost reductions as output increases are very small. As noted in the previous section, this point is known as the minimum efficient size: it is reached at 0b for industry B and 0a for industry A. We would therefore expect firms in industry A to be rather larger than in industry B. There is no further significant advantage for firms when they grow beyond these points.

Of course this minimum efficient size must be related to the size of the market. If for example industry B served a much larger market than industry A, then we would expect many more firms competing in B than in A. Some world markets have room only for a very few firms. Here, fixed costs are very high and only very large organisations can consider entry. The oil industry is an example of this.

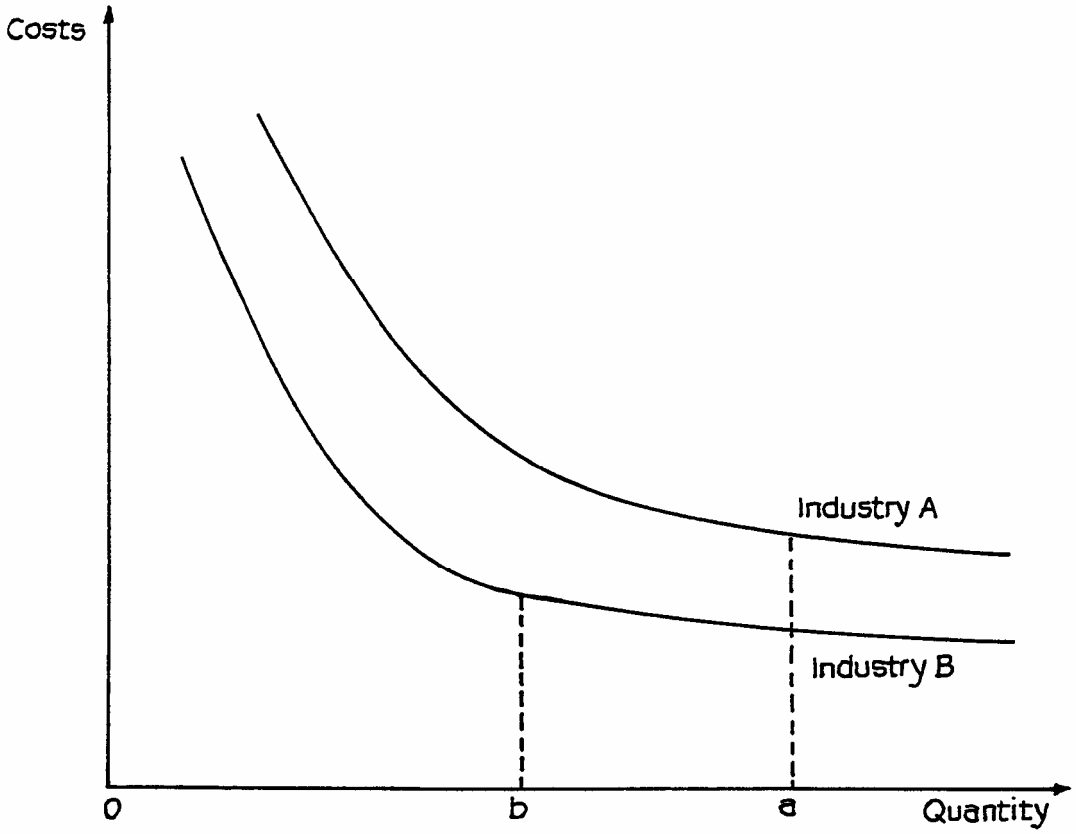


Figure 4.14: Long-run average costs for A and B

In contrast, the manufacture of many kinds of plastic household fittings does not require very expensive equipment, and many small firms are able to compete successfully in the market. The general term "economies of scale" also covers both internal and external economies, and it is only internal economies that favour large firms. External economies, such as specialised services, are available to all firms in an area or industry, and these often help small firms to survive. It is when the number of small firms drops below the level necessary for the survival of the specialist as an independent organisation that all the remaining small firms are faced with severe problems, and may have to disappear.

Special services to industry – such as industrial cleaners, photographers, and designers – often serve a restricted market and are likely to remain small. This is especially likely to be true if the service is localised. The service may only be needed occasionally by any one firm, but when it is needed the need is urgent and someone has to be found very quickly. Small local firms are better placed to provide a satisfactory service than a large national organisation.

© ABE and RRC

The MES is not the only determinant of the size of firm likely to be found within an industry. The attitudes, abilities and objectives of owners or senior executives play an important part. In the UK Marks and Spencer became a national retail chain in a period when most retail shops were small family firms, as did other high street retailers such as W H Smith, Woolworths and Boots. We can always expect to find some large firms in sectors when small firms form the majority.

At the same time we are also likely to find small firms in industries where the MES is large, apparently implying that only very large firms could survive. This may be because they serve a specialist niche which forms a small part of a larger market. Industry definitions can be misleading. For example the term "motor industry" covers activities ranging from motor vehicle assembly to the manufacture of small, specialised components. These activities are not really comparable and the MES for a component manufacturer could be much smaller than for vehicle assembly. Nevertheless it is the giant corporations which dominate the industry. If one of these fails, large numbers of the satellite firms which supply goods and services to it are also likely to fail. If the dominant firms all prosper, the satellites also flourish.

Services

Services generally tend to be smaller than manufacturing organisations, although there are, of course, some very large service firms developing in activities such as law, accounting and business consultancy. On the other hand, these large firms tend to serve large-scale customers. A leading international accountant is not really suited to do the books of the small corner shop. In any case, the shop would not be able to pay the accountant's fees. There will always therefore be small local firms of accountants, solicitors and so on. If any of these meet problems they cannot handle themselves, then they may be able to call on the specialist services of the giant.

As the service sector (including the rising leisure services) of the economy grows, so the scope for small firms continues to increase. As already suggested, new technology based on the microchip and the microcomputer/personal computer is enabling the small firm to achieve a level of administrative efficiency that would have seemed impossible only a short while ago. A business owner who can afford to spend around one to two thousand pounds on a personal computer, software packages and a printer can maintain accounting and secretarial services with just one or two people. In contrast the same standard of service would have required an office of 15 or more people 30 or so years ago – or a very expensive mainframe computer complete with specialist programmer.

The Role of Small Firms in the Economy

The part of the business sector that contains the small to medium-sized enterprises (SMEs for short), employing between 5 to 250 workers, is now recognised to be the main source of employment in most economies. Large firms tend to be visible to the public not only because of their physical size but because they usually have well-known brand names which are promoted at home and abroad by extensive marketing. But in most countries the number of very large firms is small in comparison to the very large number of SMEs. Not only do SMEs provide the main source of employment, they also turn out to be the most important source of entrepreneurial development and innovation in both products and processes in the economy. Very large companies may have large research and development (R & D) departments, and very large budgets devoted to R & D, but the evidence is that such activity is also subject to diseconomies of scale and inefficiency. In modern dynamic economies the main source of innovation tends to be the SME sector, and not the very large companies, especially the state-owned or controlled firms. The importance of SMEs for the health and growth of economies, as well as the source of most jobs, has been recognised by governments in many countries and policies have been introduced to support and promote the development of SMEs.

© ABE and RRC

Traditionally, the small-firm sector has been seen as the seedbed of enterprise and the nursery in which tomorrow's giants are reared. The microcomputer industry itself is an example. It was not the giant computer monopolists that produced the microcomputer, but brilliant electronics engineers and programmers working on their own initiative. There will always be scope for the entrepreneurial genius as evidenced by such companies as Microsoft, Apple and Google.

In recent years the earlier tendencies which resulted in large firms internalising specialised activities have been reversed. Specialist departments which had proved difficult to keep fully employed have been closed, and in many cases the specialists have been helped to form their own businesses, supported with contracts from their former employers. These newly independent firms are once again able to provide their specialist services to large and small organisations. This trend has been developed further by the growth of outsourcing and "off-shoring" of business functions to external specialist providers.

New communications technology is leading to a revival of a very old form of enterprise – what may be seen as a collection of independent firms, all working under the overall guidance of a central, largely marketing, organisation. Computer software production is often produced on this basis, with self-employed programmers producing software to detailed requirements set by the central marketing body.

Although the life expectancy of the majority of small firms continues to be short, there are nearly always people willing to fill the gaps left by the casualties. The small firm sector as such continues to exist, and the record of innovation and enterprise from small firms compares favourably with the large corporations. A healthy and dynamic economy requires a diversity of firms of all sizes and activities. Most large organisations have occasion to rely on the services of small firms: often they use them to fulfil contracts which are too small for them to carry out profitably, but which are necessary to retain the goodwill of valued customers. Moreover the continued existence of smaller rivals can often be a healthy reminder to large corporations that they are neither immortal nor indispensable. The growth of own-brand labels developed by the large supermarket chains has provided openings for many smaller manufacturers, who could not otherwise have hoped to compete with the established food corporations.

The flexibility and versatility of the modern market economy depends on the existence of many different sorts and sizes of organisation, and this diversity is essential to the maintenance of high living standards and wide employment opportunities.

© ABE and RRC

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. Explain why a firm's short-run average cost curve is U-shaped.
2. Explain why some firms' long-run average cost curve is downward sloping.
3. From the alternatives listed, complete the following:

   total cost ÷ total output =

   (a) fixed cost

   (b) marginal cost

   (c) average cost.
4. Which of the following alternatives is marginal cost is defined as:

   (a) the total cost of producing an additional unit of output

   (b) the addition to total cost from producing an additional unit of output

   (c) total variable cost divided by output

   (d) the cost saving from economies of scale as a firm increases its output?
5. Which of the following will not lead to an economy of scale as a firm expands in size:

   (a) a reduction in external costs

   (b) large-scale advertising

   (c) financial economies

   (d) transport and distribution economies?
6. A firm expands and doubles its factory size, number of employees and the number of machines and vehicles it uses in production. As a result of this increase in size its average cost of producing each unit of output falls by 20 per cent. Is this an example of:

   (a) a diseconomy of scale

   (b) the law of eventual diminishing returns

   (c) increasing returns to scale

   (d) a U-shaped short-run average cost curve?

© ABE and RRC

© ABE and RRC

Study Unit 5

Costs, Profit and Supply

Contents ***Page***

A. The Nature of Profit 76
Profit as a Factor Payment 76
Normal and Abnormal Profit 76
Profit as a Surplus 77
Summary of Explanations of Profit 78

B. Maximisation of Profit 79
Calculation 79
Profit Maximisation 83
Do Firms Maximise Profits? 84
When to Stop Producing 84

C. Influences on Supply 86
Costs and Supply 86
Supply Curve 88
Other Influences on Supply 89
Effect of Other Influences on Supply Curve 90
Relative Importance of Supply Influences 92

D. Price Elasticity of Supply 92
Calculation of Elasticity 92
Elastic and Inelastic Supply Curves 93
Elasticity of Supply in the Long Run 97

© ABE and RRC

Objectives

The aim of this unit is to: explain the concept of profit maximisation and solve problems using diagrams and data; explain the link between a firm's supply curve and its cost functions.

When you have completed this study unit you will be able to:

- explain, using appropriate examples, the difference between fixed cost and sunk cost
- explain, using words, diagrams and numerical examples, how a firm reaches its profit-maximising choice of output with reference to marginal cost and marginal revenue
- solve diagrammatic and numerical problems of profit maximisation
- explain using diagrams how a firm chooses whether or not to stay in operation or leave the industry in the short and long run
- explain how a firm's supply curve is derived from an analysis of its cost functions
- explain the reasons for movements along and shifts in supply curves
- state the formula for the elasticity of supply
- explain the effect of changes in the elasticity of supply on the diagram of a supply curve
- solve numerical problems of the elasticity of supply based on data.

A. THE NATURE OF PROFIT

The simplest definition of profit is that it is *the excess of revenue over cost*. This is a little deceptive, because in practice it is not always easy to decide what is revenue and what is cost. There are also problems arising from changes in the value of property. For example, the value of a building may rise or fall for reasons that have nothing to do with the trade carried on in that building. However at this stage it is convenient to overlook problems of this kind, and keep to the idea of profit as the excess of the revenue gained by selling products over the cost of producing those products.

Nevertheless this definition does not satisfy the economist's desire to explain why profit exists and what its economic function really is; and here we come up against two rather conflicting ideas. On the one hand there is what might be called the traditional view of profit as a payment to a factor of production, just as wage is the payment to labour or rent the payment to capital. On the other hand there is the view that profit is surplus which remains when the payments to production factors have all been made. Both views present difficulties as we shall now see.

Profit as a Factor Payment

Although considered by many as being rather old-fashioned and difficult to reconcile with modern realities, this is the view which still dominates most of the basic economics textbooks. As far as it is possible to tell, it also represents the thinking of most of today's examiners of economics in the professional examinations. You must therefore take it into account. Attempts to reconcile the idea of profit as a factor payment with the reality – that it is both very uncertain and subject to all kinds of pressures, as well as being impossible to predict or guarantee – have resulted in the development of the concepts of "normal" and "abnormal" profit.

Normal and Abnormal Profit

Here profit is seen as a payment to a fourth factor of production, the factor "enterprise". Enterprise is provided by *entrepreneurs*, people who take economic risks by organising and

© ABE and RRC

combining the other factors to produce goods and services for sale in the markets. Normal profit is thus frequently described as the reward to the entrepreneur – an attractive idea, but one which raises many difficulties.

- How do we quantify "normal"? The usual answer to this question is to suggest that it is the minimum necessary to keep the entrepreneur in the market. However, this surely depends as much on conditions in other possible markets as on the amount of profit available in the one under scrutiny. Firms that have been operating in a particular market for a lengthy period, or which operate in that market only, face greater costs of transfer to another market than newcomers, especially those which already operate in many markets. Thus, the minimum required to keep firm A in the market is unlikely to be the same amount as that sought by firm B. As economics has become more and more precise, scientific and mathematical, fewer people have been prepared to accept a concept as vague and unquantifiable as "normal" profit, in this sense.
- Who is the entrepreneur entitled to normal profit? The early economists who developed the concept were accustomed to markets containing small, individually owned and controlled firms, so that the entrepreneur who was the driving force behind the firm was usually identifiable without much trouble.

  However modern markets are dominated by large, corporate organisations with clear, bureaucratic, managerial structures. The success of this type of enterprise may lie as much in the ability of managers to reduce risks as to take them. While individual managers may be expected to show enterprise in their work, this is rarely rewarded directly with a proportionate share in profits – even if the profit attributable to the enterprise shown could be calculated. The statistical profit of the organisation belongs legally to the ordinary shareholders, who are specifically denied any right to share in management and who rarely have much detailed knowledge of the activities of the organisation. When we further recognise that modern large public companies are likely to operate in many markets in many countries, we have to agree that all this is impossible to reconcile with the definition of normal profit.

However if it is accepted that there is such a thing as normal profit then this implies that there can be "abnormal" profit. Some textbooks do in fact describe all profits above the normal as abnormal. Others, clearly unhappy at the emotive implications of this term, use the less derogatory "supernormal". In either case, the impression is usually given that firms should not be permitted to earn profits above normal.

Instead of either abnormal or supernormal, some writers have referred to what they call "pure profit", by which they appear to mean any surplus over and above all payments to factors including the normal profit due to the entrepreneur.

Profit as a Surplus

If we see profit not as a factor payment but as a surplus remaining after the production factors have been paid for, the question then arises as to who owns, or should own, this surplus.

To Marxist economists the answer is clear. Economic value is created by human labour, without which there can be no economic activity. The berries growing wild on the bush belong to the picker, whose labour of picking has turned them into food. Thus any surplus created by work belongs to those who carry out the work. Therefore profit, to the Marxist who does not recognise a separate entrepreneur, belongs to the workers. However, the Marxist recognises that in the modern capitalist society where production is organised by the owners of capital and, in the Marxist view, for the benefit of the owners of capital, profit, is in practice, allocated to the owners of capital.

If this view is accepted, profit, not interest, becomes the payment to the owners of capital. To the Marxist, the fact that it is paid to the owners of capital rather than to the rightful owners,

© ABE and RRC

the contributors of labour, is the result of the domination of capital over labour in the modern capitalist society.

In support of this view it is possible to point to company law, which provides that a company's profit belongs to the company's shareholders or, more precisely, to the contributors of the "risk capital" or "equity", the ordinary shareholders – in American terminology, the common stockholders. There is no legal requirement that the company should share its profits with the suppliers of labour (employees) or with the suppliers of loan capital, who receive their agreed rate of interest.

Still largely accepting this concept of profit as a surplus, other economists, some of whom belong to what has been called the "Austrian school", take a very different view of its economic function. They see it as the driving force of the modern economy and the incentive which has been largely instrumental in bringing about the enormous improvement in general living standards in the market economies over the past two centuries. They see the striving for profit as the force that produces new products, new production technology, new forms of business organisation and new uses for basic resources. The profit that produces this economic energy and invites people of all kinds to take risks with their own resources of money, time and futures, is not "normal profit" but the largest possible profit that can be made in the circumstances within which business operates. There is no need to distinguish between normal and abnormal profit. All profit is necessary to stimulate future economic activity and to provide the investment finance necessary to make the activity possible and raise the level of technology.

Unlike Marxists, the economists who take this view do not see profit as being stolen from workers, nor do they see any need for labour to be given only the lowest possible wage. Indeed for business enterprise to succeed, goods and services have to be sold to workers whose incomes are well above subsistence levels, who have disposable incomes and the freedom to choose how to spend these incomes and who expect to have rising incomes. Workers therefore benefit from profitable economic activity by earning rising wages.

Summary of Explanations of Profit

One economist who recognised the various ways in which profit has been explained was the great American writer and teacher, Professor Samuelson. He identified six distinct "views", which can be summarised as follows:

(a) Profit is seen as a *balancing item* and a result of accounting conventions but should properly be seen as a return to one or more of the production factors. For example, most of what accountants show as the "profit" of the majority of small family firms would better be described as the proprietor's wage for his or her physical and mental effort and interest on his or her personal savings invested in the business.

(b) The second view sees profit as a *reward to "enterprise and innovation"* and a return for the temporary monopoly achieved by being first in the field with a successful new commercial idea.

(c) The third sees profit as a *reward for successful risk-taking.* Although willingness to take risks does not always (or often) bring compensating profits, it is usually the hope of earning such profits that provides the spur to help business people overcome their natural inclination to avoid risk.

(d) The fourth view simply takes the third view further; profit is a *positive incentive to "coax out the supply of risk-bearing capital"*. It is the high return sought by providers of what is often known as "venture capital".

(e) The fifth view regards profit as a *return to monopoly*, whether natural or achieved by artificial means. It is this association of abnormal profit with monopoly that has coloured so much teaching about business profits and objectives.

© ABE and RRC

(f) The sixth view recognises the Marxist explanation of profit as *surplus value* which, for Marx, was properly the reward of the labour that created the value but which, in a capitalist economy, is appropriated by the owners of capital.

Clearly there is no simple or generally agreed explanation of the economic function of profit, though most would agree that both profit and a spirit of enterprise are extremely important elements in modern market economies.

B. MAXIMISATION OF PROFIT

Calculation

We can arrive at the amount of profit for any given level of output in at least two ways. We can calculate total revenue and total cost and find the difference, or we can calculate the average revenue and the average cost, find the difference and multiply this by the quantity sold.

We shall first consider profit as the difference between total revenue and total cost. Suppose we return to the example of the last study unit and assume that all units of the product are sold at a given market price of £210 per unit. Costs remain as before. We can now show total revenue and cost columns for each range of output up to 150 units per week – as in Table 5.1.

Quantity	**Total Cost**	**Total Revenue** (output level × £210)
(units per week)	*£*	*£*
0	10,000	0
10	11,000	2,100
20	11,600	4,200
30	12,000	6,300
40	13,000	8,400
50	14,000	10,500
60	15,000	12,600
70	16,000	14,700
80	17,000	16,800
90	18,150	18,900
100	19,500	21,000
110	21,150	23,100
120	23,250	25,200
130	26,000	27,300
140	29,550	29,400
150	34,000	31,500

Table 5.1: Total cost and total revenue

From this table we can see that revenue exceeds total cost at output levels 90 to 130 units per week. At all other output levels, total costs are greater than total revenue, so losses would be suffered.

Table 5.2 shows the profit at each output level.

© ABE and RRC

Quantity	Profit £
90	750
100	1,500
110	1,950
120	1,950
130	1,300

Table 5.2: Profit at different output levels

The position is illustrated in Figure 5.1, where the shaded area represents the profit produced when total revenue is greater than total cost.

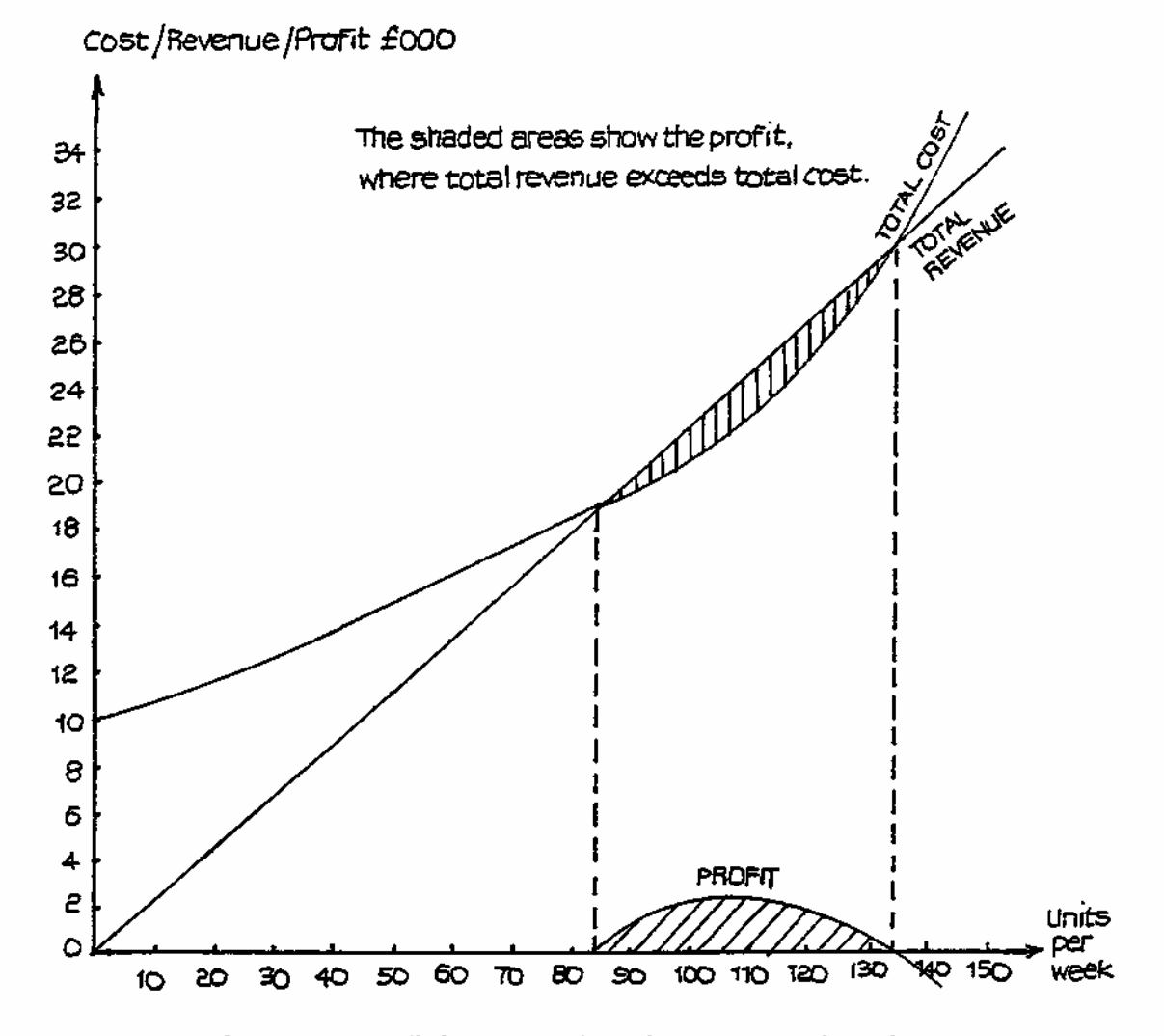


Figure 5.1: Profit in terms of total revenue and total cost

The same position is shown by the average cost and price/average revenue curves of Figure 5.2. In this case however the shaded area does not represent the total profit, but the profit per unit of output. Total profit would be given by multiplying the profit per unit by the number of units produced.

In this example, the firm is selling all units at a given price, so that the total revenue curve continues to increase – though this does not of course mean that it is possible to make a profit at output levels above 130 or so units per week.

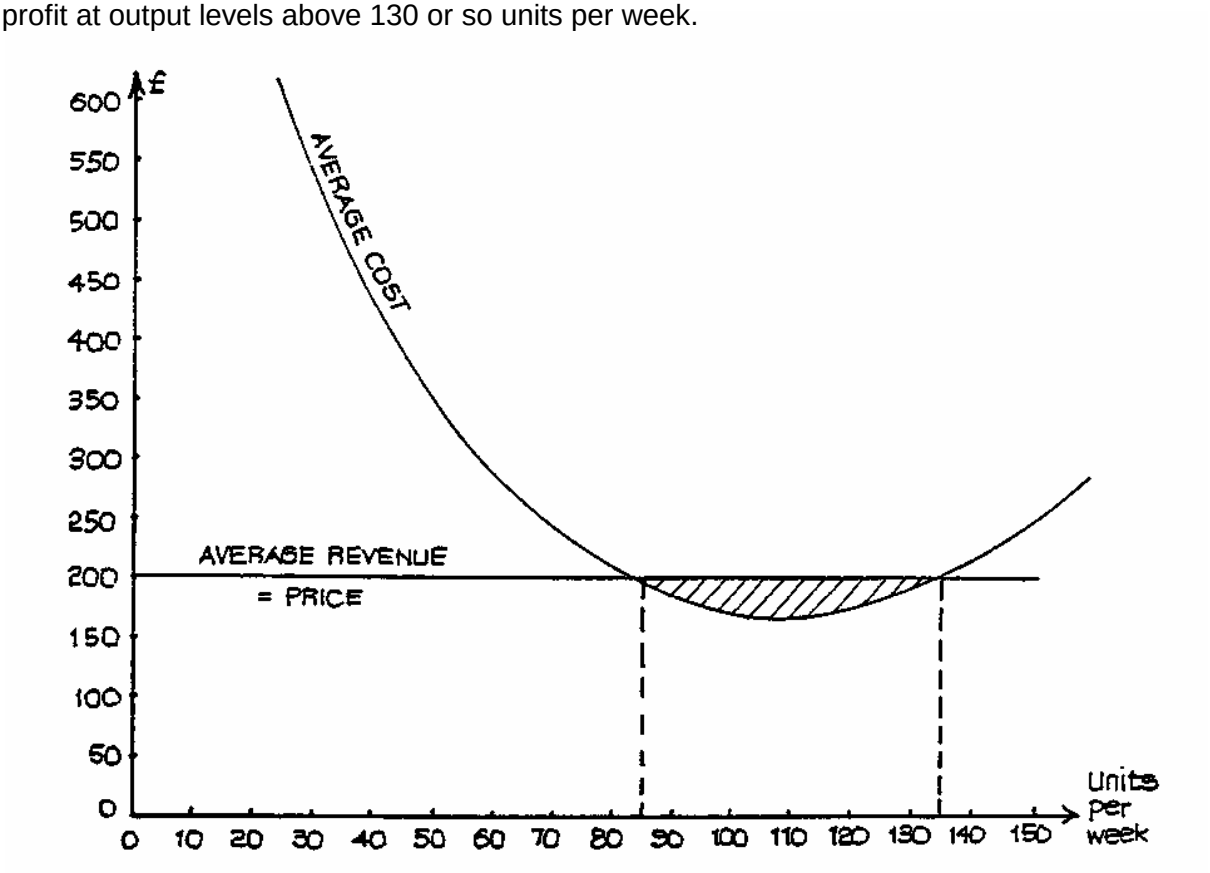


Figure 5.2: Profit in terms of average revenue and average cost

We saw in an earlier study unit that the revenue position could be rather different where the firm had to reduce price in order to increase output. Such a situation is illustrated in Figure 5.3. No specific figures are shown here – this is a general model – and it shows that the firm can make profits at all output levels between Oa and Ob.

These levels, where total revenue just equals total cost, are called the break-even output levels or sometimes break-even points.

It is often more convenient to show the average cost and revenue curves (see Figure 5.4).

If we assume that the firm is selling all units at any given output level at the same price (i.e. is not discriminating between different customers over price) then the average revenue curve is also the price/output curve (i.e. the demand curve). In this model, we can also see that the firm makes profits between output levels Oa and Ob. This is the quantity range where average revenue is greater than average cost.

© ABE and RRC

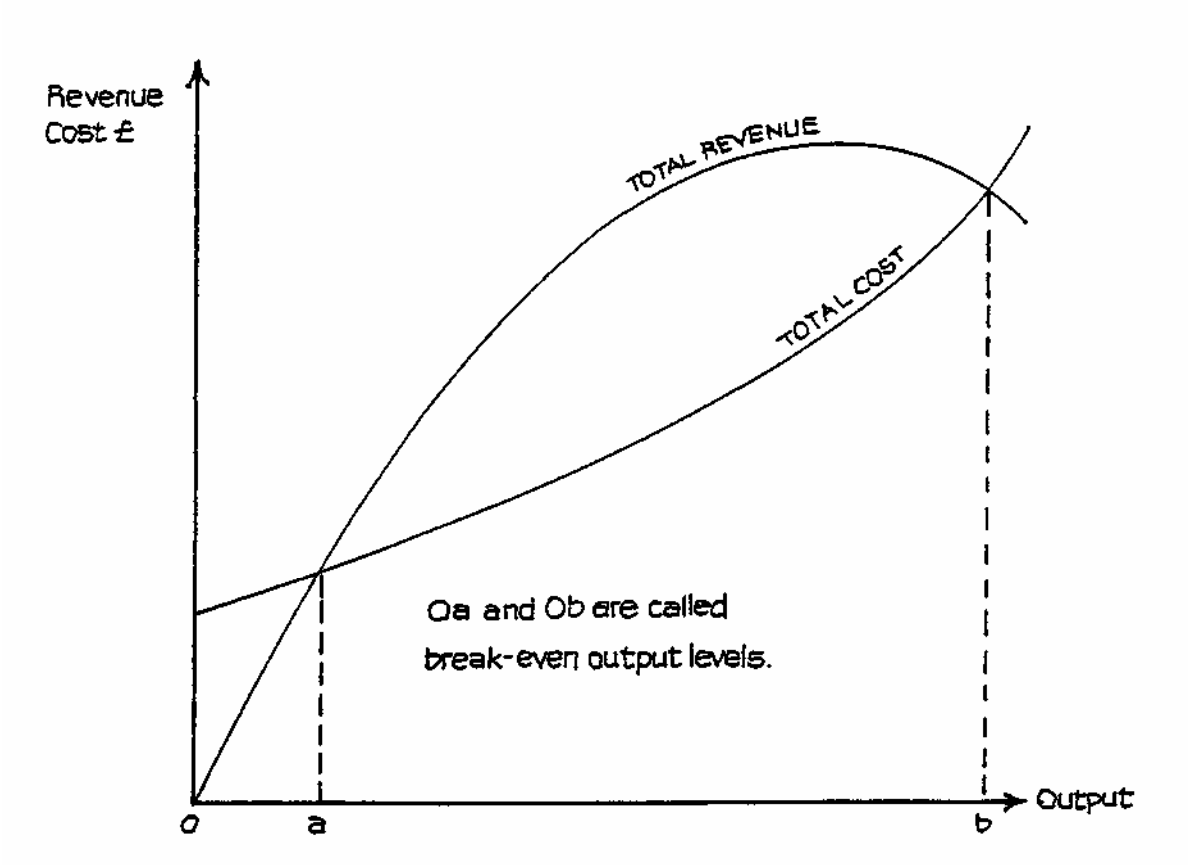


Figure 5.3: Break-even output levels

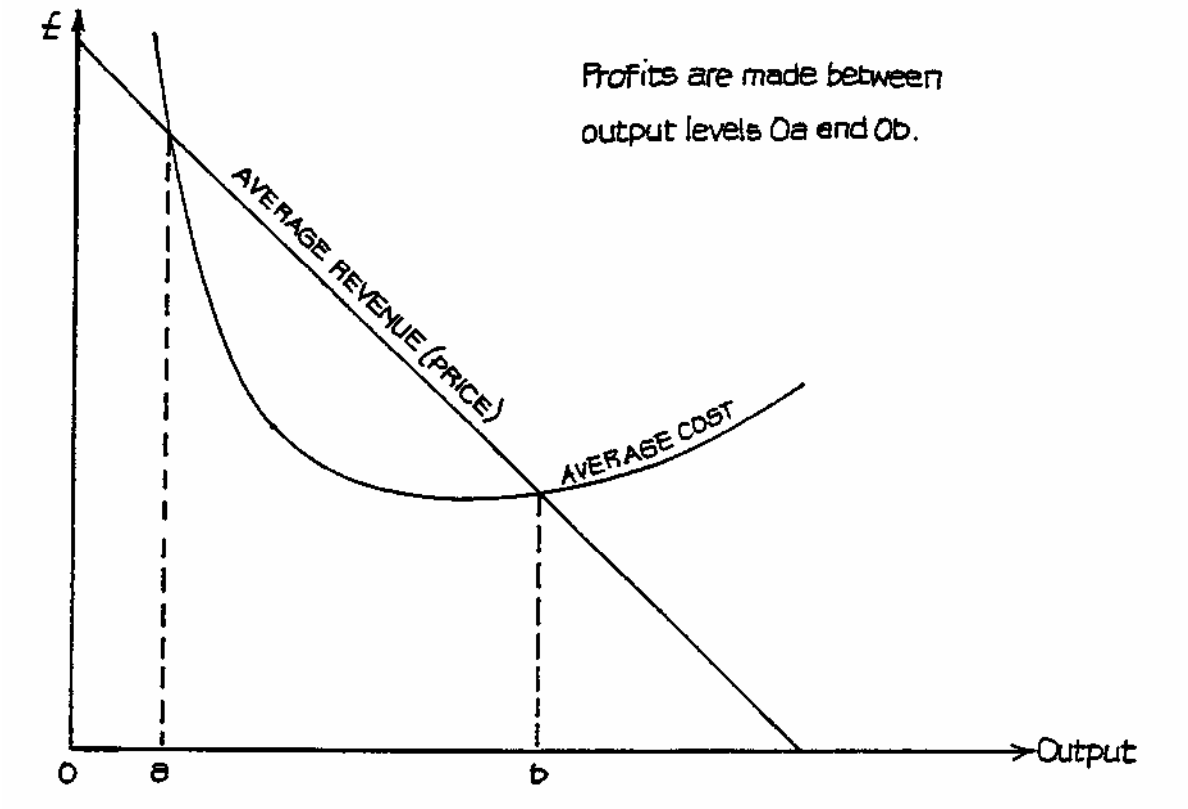


Figure 5.4: Profits, average cost and average revenue

Profit Maximisation

So far we have seen the output levels where profits are made, but we have not yet identified the output level where the largest possible (maximum) profit can be made. However, if we refer back to our profit table, we see that there are two points where profits are at their largest – at output levels of 110 and 120 units per week. Here, total profit stays at £1,950. If the firm wants to make the largest possible profit, it can choose either of these two levels. It is not unusual for profit to have a rather "flat top" and stretch across two stages in this way. In other cases it can peak at a single stage.

Now look back at Table 4.4 in the Study Unit 4, which showed marginal costs. Bearing in mind that we assumed the firm to be selling at a constant price of £210, look at the marginal cost column. We have explained that, when the firm can sell at a constant price at all levels of output, the price is also the average revenue and the marginal revenue. Thus, in this case, the firm's marginal revenue is £210. If you look down column 4, you will see that the marginal cost is £210 at the midpoint, representing the change from output level 110 to 120 units per week. This is precisely the output range where profits are at their highest level, i.e. £1,950.

This is no accident. It illustrates the general rule that profits are always maximised at the output levels where marginal cost is equal to marginal revenue.

The general position is illustrated in Figures 5.5 and 5.6. Figure 5.5 shows the case where average revenue equals marginal revenue (constant price at all output levels) and Figure 5.6 shows the sloping average revenue curve with the marginal revenue curve in the correct position, as we explained before.

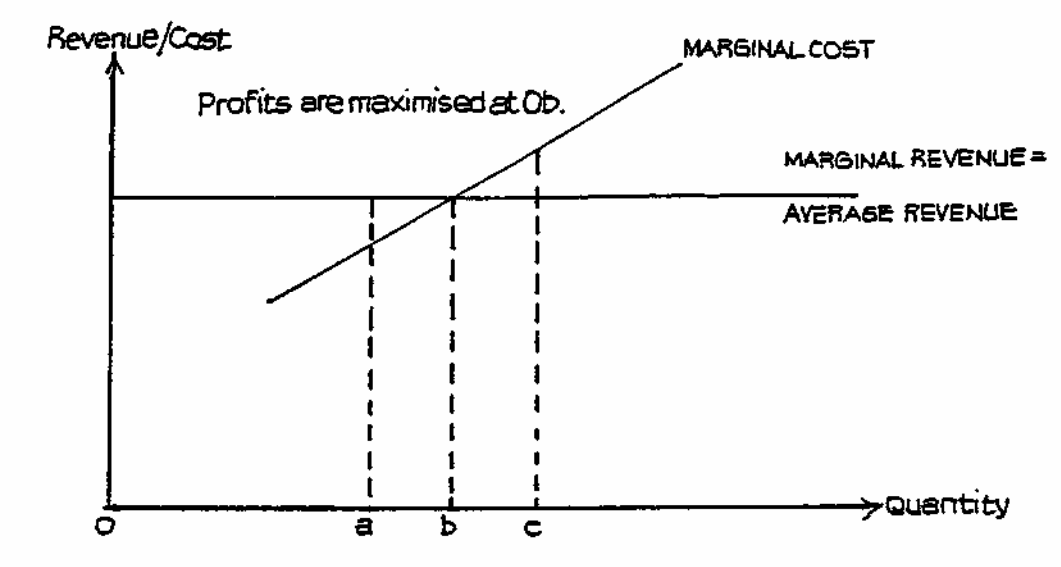


Figure 5.5: Profit maximisation – marginal revenue equals average revenue

In both cases, the argument is the same. It does not matter whether the marginal revenue curve slopes or not. If the firm produces at output level Oa, i.e. below the level where marginal cost equals marginal revenue, it would pay it to increase output because the revenue received for each additional unit is greater than the cost of producing that unit. If the firm is producing at output level Oc, above the level where marginal cost equals marginal revenue, then it will pay it to reduce output because revenue lost for each unit of output sacrificed is less than the cost of its production. Only at output level Ob, where marginal cost equals marginal revenue, will it pay the firm to stay at the same level. It cannot then increase profit by any change in quantity produced. This is the level where profits are maximised.

© ABE and RRC

This is a most important rule which you should remember carefully, i.e. *to maximise profits the firm produces at the output level where marginal cost is equal to marginal revenue*.

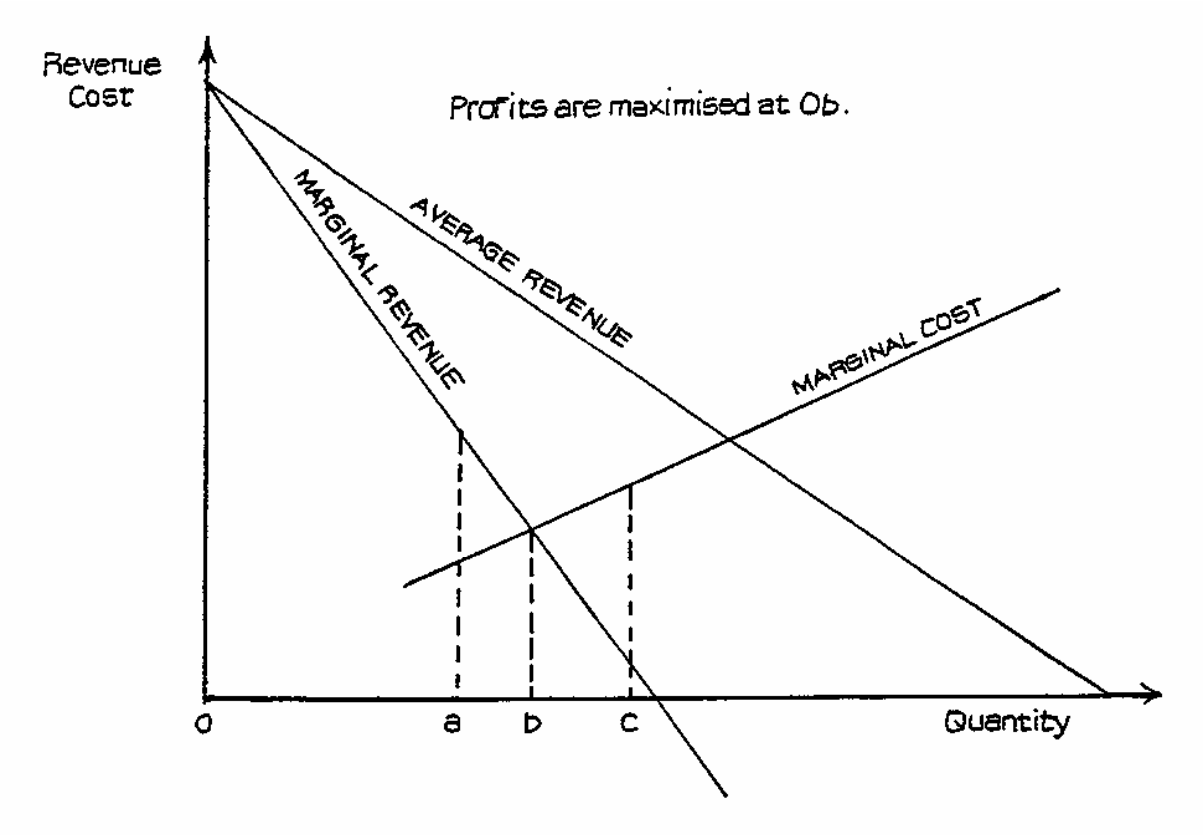


Figure 5.6: Profit maximisation

Do Firms Maximise Profits?

It is often argued that we should not automatically assume firms do seek to maximise profit. It is suggested that they may have other objectives, e.g. to maximise revenue, to increase output or to achieve a given share of the market, or simply to please and reconcile the conflicting objectives of shareholders, managers and employees.

All this may be true – many firms may not be seeking to maximise their profits. Many may not have sufficient information about market demand and their costs to maximise profits even if they wished. On the other hand, this does not rule out our view that the profit-maximising output level and the rule for achieving this are matters of very great importance for an understanding of business decisions. The firm may decide to sacrifice some profit in order to pursue some other objective, but it should know how much profit is being sacrificed.

An assumption of profit-maximising behaviour is an essential starting point for the analysis of the business organisation. As long as we recognise that it is not necessarily the finishing point, then we can accept this assumption at this stage of our studies unless there is a very good reason to do otherwise.

When to Stop Producing

Firms are in business to make a profit. What should a firm do if it cannot make any profit? When should a firm close down and leave the market?

The answer to these questions is straightforward for the longer-term period. If a firm cannot cover all its costs and operates at a loss it will quickly become insolvent and cease production. But should a firm always cease production if it runs at a loss? The answer is not in some circumstances.

© ABE and RRC

There are conditions in the short run when a firm should continue to produce, despite not being able to cover all its costs. This is because if it were to cease production its loss would be even greater. To understand how this can happen it is necessary to return to a consideration of a firm's costs. A firm's total cost of production consists of two components, fixed costs and variable costs. Variable costs are the wages of staff, the cost of the materials used in production and the cost of energy, such as electricity or fuel oil. Clearly, if a firm stops production it no longer needs such variable inputs and can immediately reduce its costs accordingly. The same is not true for the firm's fixed costs.

Fixed costs can include such things as an annual property tax or business rate on a firm's factory or offices, the annual rent paid to the owner of the buildings or land used by the firm, contracts to hire machinery or vehicles, and even annual employment and salary contracts for some of the senior or technical staff. All of these costs have one thing in common: they are agreed or known in advance. Contracts are signed and require payments to be made for an agreed period which could be months or several years. If the firm ceases production it is still contractually obligated to go on paying these fixed costs, unless the terms of agreement allow it to cancel its contracts, or the contracts come up for renewal. Thus in the short run a firm is faced with costs even if it produces nothing. This fact has an important implication for the firm's decision to cancel or continue production in the short run, even when it knows that it will stop producing in the long run. *Provided a firm can cover its variable costs of production and make some contribution to its fixed costs it should continue to produce in the short run.* By continuing to produce the contribution it makes to its fixed costs it reduces the magnitude of its loss. That is, if a loss is unavoidable in the short run, a smaller loss is preferable to a larger loss. Despite the fact that it involves a loss this is actually another example of profit maximising behaviour in the sense that the firm is minimising its loss which is the best thing it can do in the situation it faces.

Figure 5.7 illustrates the logic of such a decision. In the graph the firm's average fixed cost curve falls continuously from left to right, because as it increases production its fixed costs are spread over more and more units of output and become less and less significant. The firm's average variable cost curve has the usual U-shape, reflecting the law of eventual diminishing returns. The firm's average total cost curve is the sum of its average fixed cost and average variable cost. Because average fixed cost becomes smaller and smaller as output increases the average total cost and average variable costs curves move closer and closer together at higher levels of output. The firm's marginal cost curve is also shown in the graph. To determine the firm's profit maximising level of output we also need to know its marginal revenue curve. Suppose, for ease of exposition, that the firm is operating in a market situation where it can sell every unit of output at the same price. In this case its marginal revenue curve is a horizontal straight line at the level of the market price. It is also its average revenue curve.

In Figure 5.7 the profit maximising point where MC equals MR occurs at a price which is below the firm's average total cost. If its average revenue is less than its average cost it also follows that its total revenue must be less than its total cost and production is making a loss. Nevertheless, it still makes sense for the firm to continue to produce output OQe in the short-run, despite its loss, because at that output level it is covering its average variable costs and part of its fixed costs. Its optimum output is OQe because it minimises its loss in the short run. In the longer run all costs are variable and the firm will cease production unless the market price increases to a level at which its total revenue exceeds its total costs.

© ABE and RRC

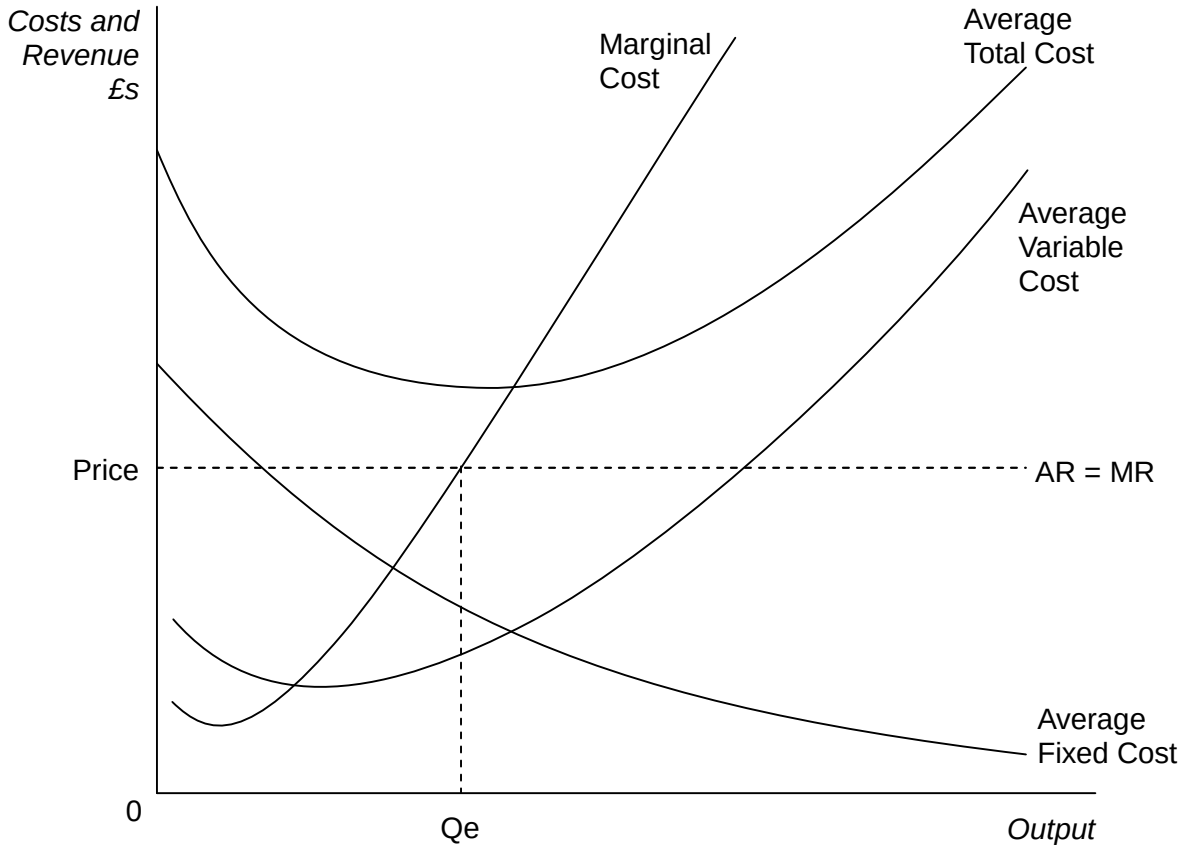


Figure 5.7: Loss-making production

We can now derive a decision rule for firms regarding whether they should continue or cease production in the short run even when production is unprofitable. A firm should continue to operate at a loss in the short term provided its average revenue exceeds its average variable cost. That is, by choosing to produce anywhere in the range between its average variable cost and its average total cost, the difference between them being average fixed cost, the firm is recovering some of its fixed costs and reducing the magnitude of its unavoidable loss in the short run.

C. INFLUENCES ON SUPPLY

Costs and Supply

If we accept that business firms exist to make profits, then we can recognise that there must be a close link between costs, profits and the willingness of firms to produce the goods and services that consumers wish to buy. After all, profit is the difference between revenue and costs, so that at any given price the amount of profit will depend on production costs. If price remains constant and costs rise, then profit falls and we can expect firms to be less willing to supply goods and services. Similarly, if costs remain unchanged and price rises, then profits will rise and firms will wish to supply more in order to secure the increased profit.

We thus have no difficulty in accepting the link between costs and the amount that firms are prepared to supply at a given price or range of prices. If we accept the aim of profit maximisation, then we can be a little more precise than this.

Suppose a firm is seeking to maximise profits and can sell all it can produce at the ruling market price. Suppose too that this market price can change. What will then be the firm's

© ABE and RRC

response? Look at Figure 5.8. The profit-maximising firm will seek to produce at that output level where marginal cost is equal to price, i.e. at quantity Oq at price Op, at Oq_1 at price Op_1, and Oq_2 at price Op_2.

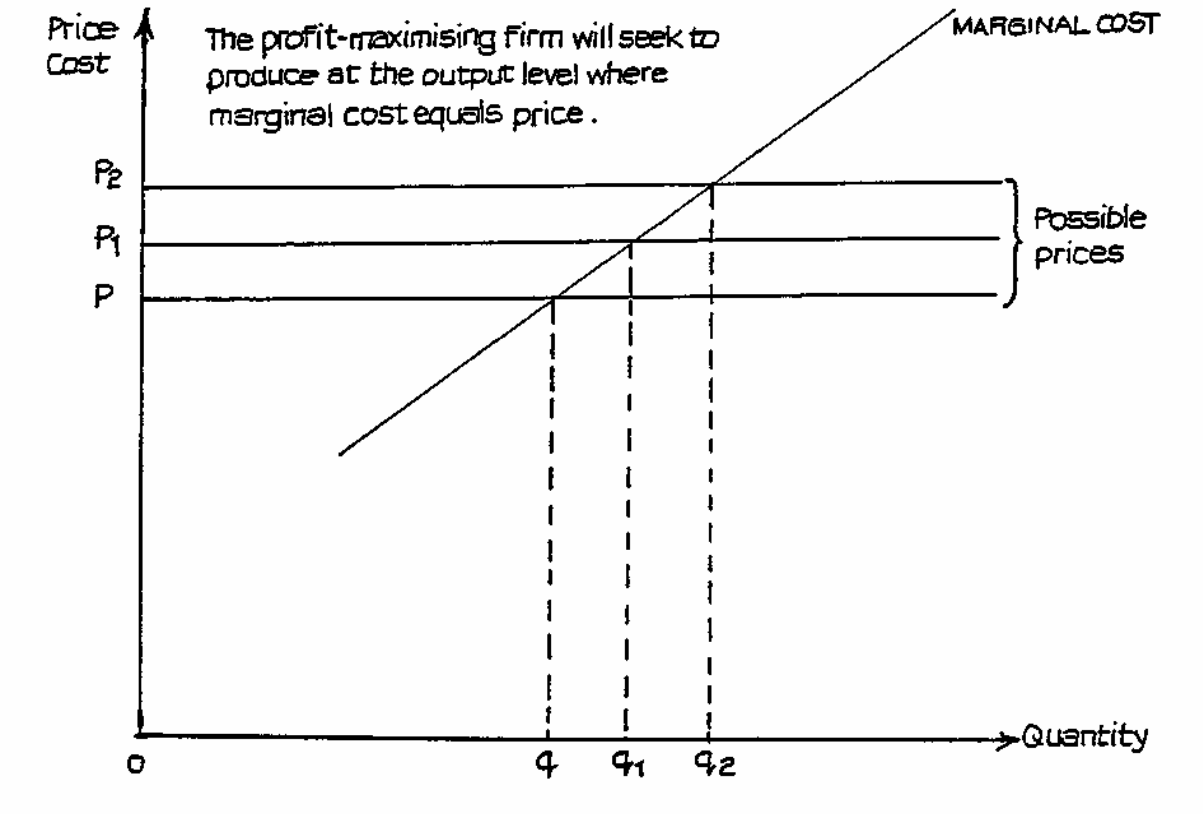


Figure 5.8: Profit-maximising output levels

Thus we can see that the firm will increase the quantity it is willing to supply as price increases – and, conversely, reduce quantity as price falls – and that the actual change in quantity will be governed by the marginal cost curve.

Therefore under conditions of perfect competition, the individual firm's supply curve is its marginal cost curve. Consequently, the market supply curve is derived from the sum of the marginal cost curves of all the firms operating within the market.

This argument continues to hold good when we abandon the assumption of the firm accepting the market price. If a firm faces a downward-sloping demand curve for its product, and hence a downward-sloping marginal revenue curve, we still get the same increase in quantity following the marginal cost curve if we again move the marginal revenue curve outwards, further from the point of origin. This is shown in Figure 5.9.

Notice though that Figure 5.9 is drawn on the assumption that the average revenue curve is moving outwards evenly and with its slope unchanged. There is no guarantee that this will ever happen in practice. If the slope of the average revenue curve changes, then so too will the slope of the marginal revenue curve, and there will no longer be the smooth increase in quantity suggested by Figure 5.9. For this reason, we cannot say that, in imperfect markets, the market supply curve will represent the sum of the marginal cost curves of the individual firms. Nevertheless, the general link between supply and marginal costs remains, although it is unlikely to be as direct as in perfect competition.

© ABE and RRC

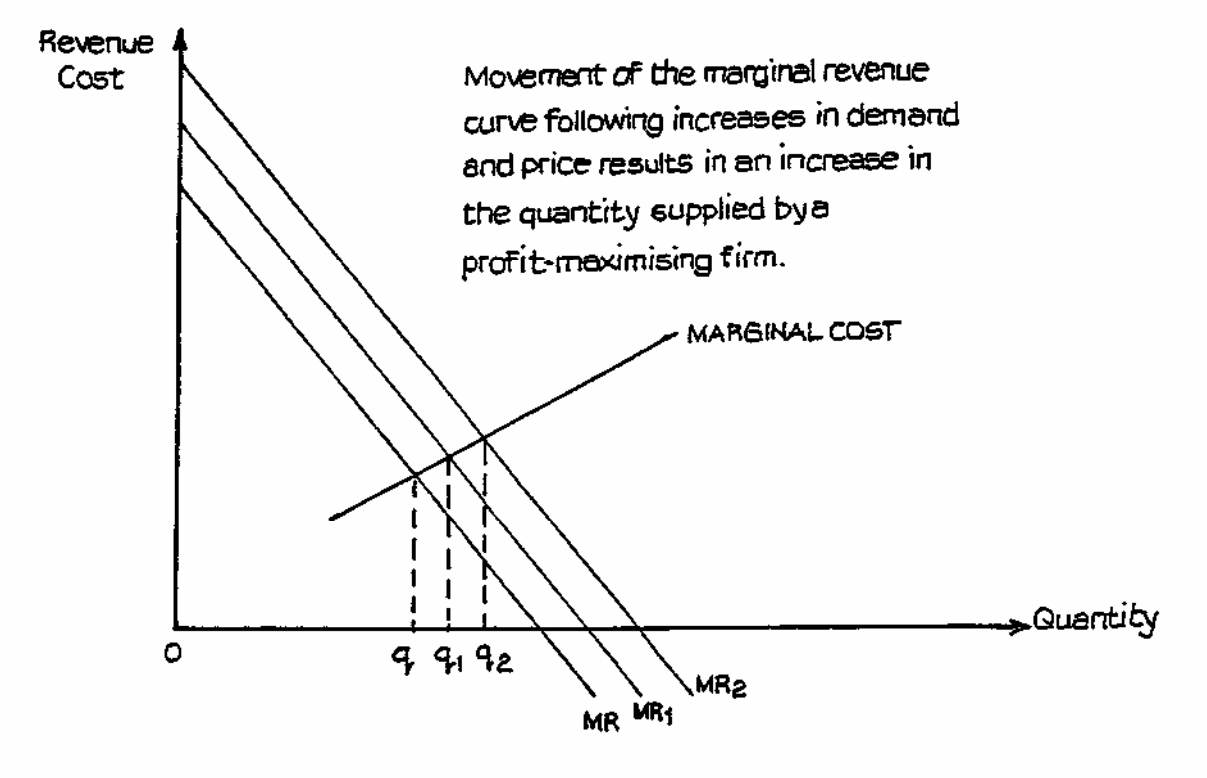


Figure 5.9: Movement of marginal revenue curve

Here again, a movement of the marginal revenue curve produces a shift in quantity supplied, in accordance with the marginal cost curve.

If you wish you can add the average revenue curves to this graph, and thus show the prices corresponding to the three quantity levels Oq, Oq_1 and Oq_2. Remember the relationship between average and marginal revenue, and remember that price will be shown by the vertical line from any given quantity level to the average revenue curve.

Supply Curve

If we accept the view that firms will seek to increase the quantity supplied if price increases, and reduce it if price falls, then we can produce a supply curve showing the amounts involved. A supply curve can be for an individual firm – in which case, assuming profit-maximising objectives, it will be the marginal cost curve – or for all firms supplying a particular product, where it will be made up of the sum of the marginal cost curves of all the firms supplying the product.

However the supply curve is formed, we can accept that its general shape will be as in Figure 5.10. This shows the general assumption that more will be supplied as the price rises – all other influences remaining the same.

© ABE and RRC

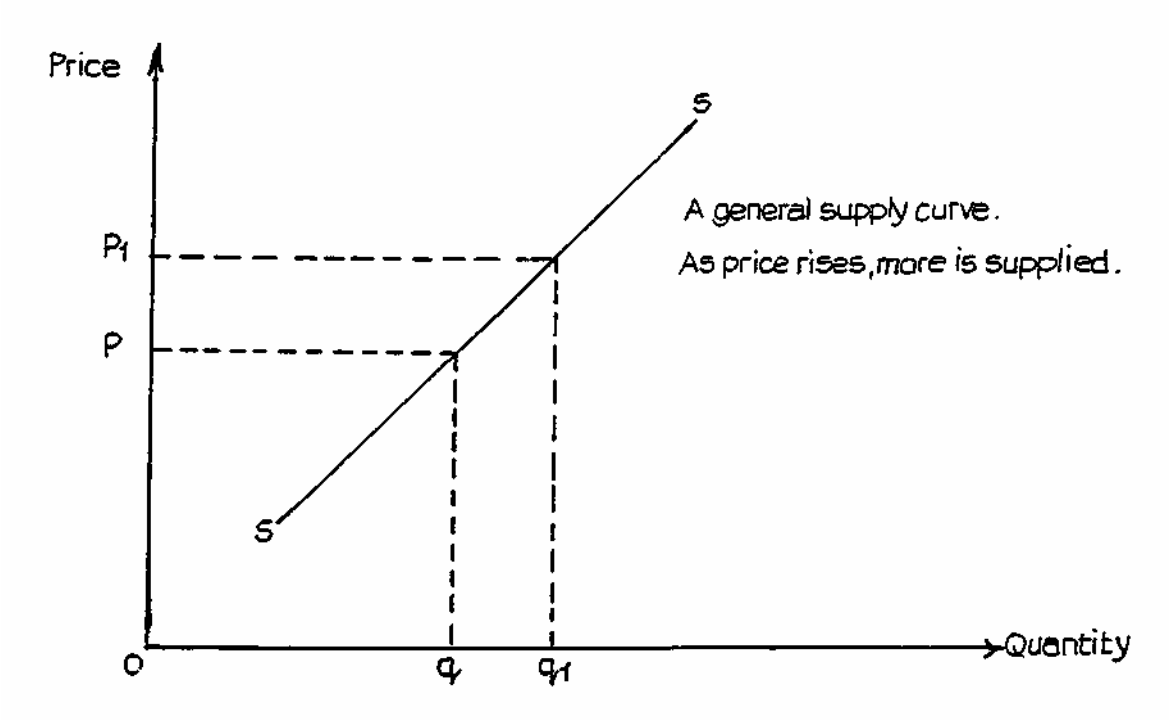


Figure 5.10: A general supply curve

Other Influences on Supply

The concept of the supply curve reflects the view that price is one of the most important influences on the quantity supplied. However there are other influences, and these are mostly concerned with the cost of production and with profits. Remember that in a market economy, the great driving force for supply is profit, so anything that affects profit will affect supply. In very broad terms, since profit is the difference between revenue and costs, supply will be directly affected by anything affecting revenue, price and costs.

We can summarise some of the most important influences as follows:

- **Costs of Factors and Other Inputs**

  Any change in costs, with price staying constant, will change the profit expectations and will thus influence decisions regarding supply. For the profit-maximising firm, a change in variable costs will change the marginal cost curve, and so change the supply schedule. Examples of factor costs include wages, land and property rents, interest rates on capital, basic material prices and the prices of fuel and power. Any of these may also affect the prices of intermediate products and services required by the firm, and so further influence supply.

- **Changes in Taxes**

  If a government tax is charged at any stage of production or on the profits of the business, then any change in the tax rate will affect the profits anticipated from supply, and thus affect supply intentions. An increase in a production tax, such as value added tax, will have the same effect as an increase in factor costs; it will tend to reduce the quantity that firms are willing to supply at all prices in a given range.

- **Changes in Technology**

  By technology is meant the methods of combining factors and inputs in order to achieve production. An improvement in technology, which allows a given level of production to be achieved with fewer factor inputs or with a different combination of

© ABE and RRC

factors, so that the total cost is lower, will tend to increase the quantity likely to be supplied at all prices within the range. Some types of technology may be possible only if production is required on a large scale. This can have a marked effect on supply. Thus, small-scale supply may be possible only at much higher prices than large-scale supply, when the different technology becomes worthwhile. The result may be to shift the whole supply curve when production reaches the critical level required for the large-scale technology.

- **Efficiency of the Firm**

  Multinational production of similar products has shown that firms in country A can sometimes produce more from a given combination of labour and capital than similar firms in country B, even though production methods and levels of technology are all much the same. Differences in the productivity of labour and capital (the amount produced per unit of labour and capital) must, in these cases, be caused by differences in managerial efficiency or in the conditions under which people work. In some cases, the movement of managers from one country to the other makes little difference to the gap in factor productivity. The causes of these differing levels of efficiency are not fully understood, but they do help to explain why large multinational firms tend to prefer some countries to others. A change in the level of business efficiency will of course influence supply.

- **Changes in Relative Profitability of Products**

  If a firm can produce either product X or product Y from similar factors, machines and skills, and if it becomes more profitable to produce Y, then the firm is likely to switch its production activities from X to Y. This may happen if the firm normally makes X, but the price of Y rises while the price of X stays the same.

  There can be other causes of production switches. If there are numbers of firms able to choose between producing X or Y, and the market for Y suddenly disappears, perhaps because of a political decision, then firms previously making Y will have to switch to X if they wish to remain in business. The result will be to increase the supply of X at all prices.

Effect of Other Influences on Supply Curve

All these changes can be illustrated by a movement of the whole supply curve, indicating a change in supply intentions throughout the given price range. Such a shift in the supply curve is illustrated in the general graphical model of Figure 5.11.

© ABE and RRC

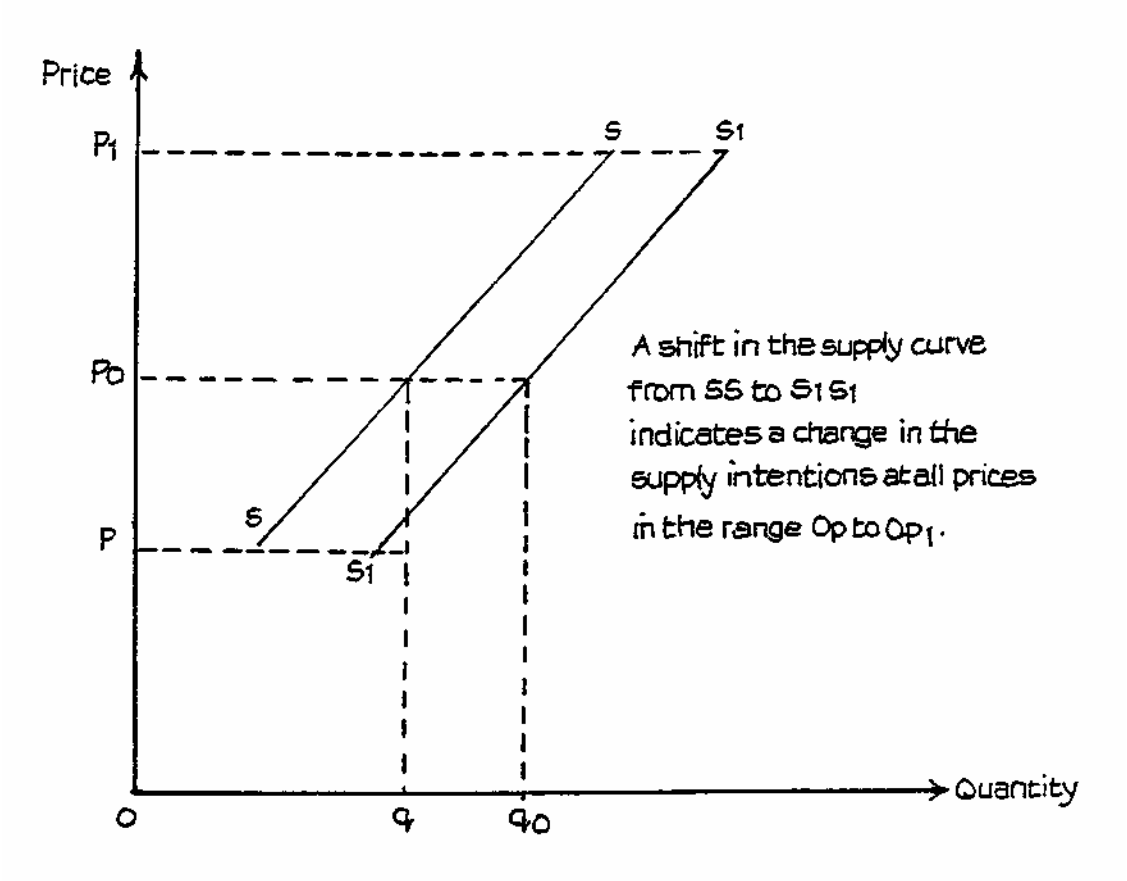


Figure 5.11: A shift in the supply curve

A shift of this type may follow a change in one or more of the influences as previously described. Moreover, several influences may be operating in different directions. For example, a tax increase may be depressing supply intentions while an improvement in technology is raising them. The final result depends on the relative strength of the influences. It is not easy to analyse these effects through simple graphical models. This is why more advanced studies make rather more use of algebraic models which can be easily handled by computers, and why you should begin to become familiar with functional expressions such as the following.

$Qs = f(P, C, T, v, y, \pi_o)$

where:

Qs = quantity of a product supplied

P = product's price

C = factory and input costs

T = business taxes

v = level of technology

y = level of business efficiency

π_o = relative profitability of products.

This simply states that quantity supplied is a function of, or is dependent on, the various influences symbolised.

© ABE and RRC

Relative Importance of Supply Influences

As with demand, different products will be affected to different degrees by the various influences on supply. In the case of supply, much will depend on the methods of production and the ease with which producers can respond to changes in factor costs and availability as well as in technology. Consequently, it is easier to assess the relative importance of the influences on supply than those on demand. A careful study of production technology and relative factor costs will indicate which are likely to have the most impact on producer intentions. A production process heavily dependent on labour (labour-intensive) will be more responsive to changes in wage levels than one that is highly mechanised or automated and thus capital-intensive. On the other hand, production which is highly capital-intensive will be more vulnerable to changes in interest rates, since much capital is likely to be borrowed in one form or another. The potential costs of changing production levels tend to be greater with capital-intensive production methods.

D. PRICE ELASTICITY OF SUPPLY

Calculation of Elasticity

The concept of elasticity, which we applied to demand, can also be applied to supply. However, here it is usually only price elasticity with which we are concerned. The method of calculating supply elasticity is exactly the same as for price elasticity of demand, i.e.

$$\text{supply elasticity of a product } (E_s) = \frac{\text{proportional change in quantity supplied}}{\text{proportional change in the product's price}}$$

or

$$E_s = \frac{\Delta Q_s}{Q_s} \div \frac{\Delta P}{P} = \frac{P \Delta Q_s}{Q_s \Delta P}$$

Notice that the value of E_s is always positive (i.e. greater than zero). This is because the change of quantity is in the same direction as the change in price.

Figure 5.12 shows an example of a simple supply elasticity calculation.

Notice here that figures for both P and Q are obtained from the midpoint of the change in price and quantity, so that the calculation is the same for both a rise and a fall in price. Notice also that the result of this particular calculation is that E_s equals unity (1).

If you calculate values for E_s at any other price level on this curve, you should obtain the same results. The reason for this is explained shortly.

© ABE and RRC

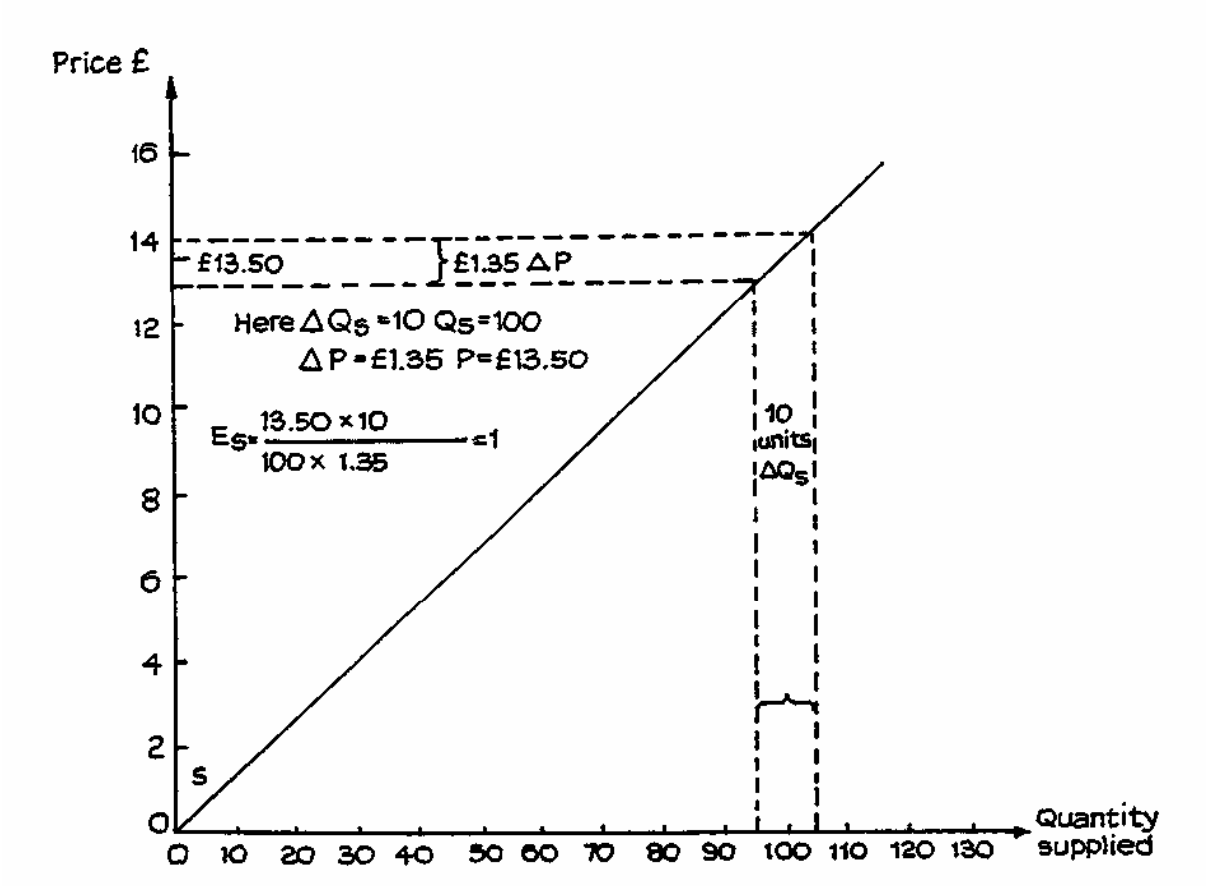


Figure 5.12: Supply elasticity calculation

Elastic and Inelastic Supply Curves

Price elasticity of demand was shown to change as price changed. A rather different position arises in the case of supply elasticity. We said that the value of E_s for the supply curve of Figure 5.12 would always be 1. This is because the curve starts at the point of origin.

A simple proof follows, relating to Figure 5.13. The proof assumes a knowledge of simple geometry.

© ABE and RRC

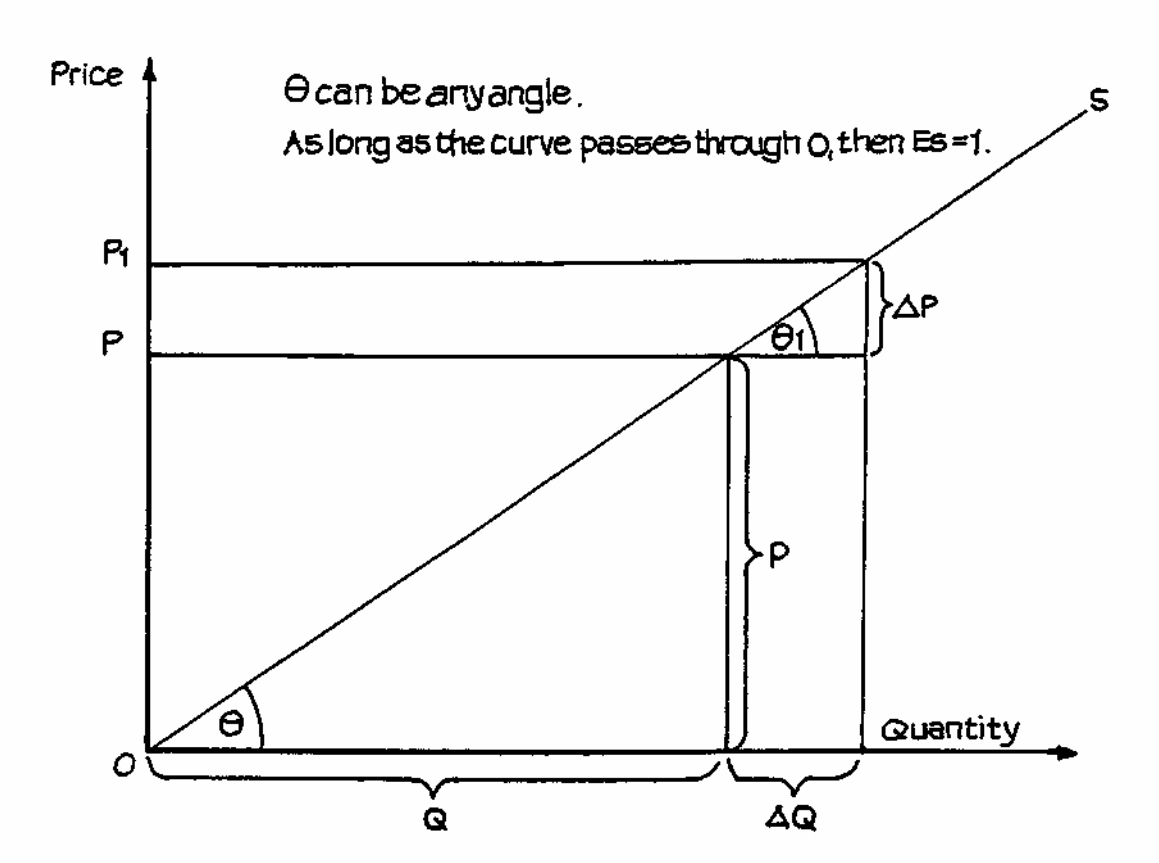


Figure 5.13: Proof of $E_s = 1$

From the diagram in Figure 5.13:

$\theta = \theta_1$,

$\frac{P}{Q} = \tan\theta$ and

$\frac{\Delta P}{\Delta Q} = \tan\theta_1$

so,

$$\frac{P}{Q} = \frac{\Delta P}{\Delta Q}$$

But,

$$E_s = \frac{\Delta Q}{Q} \div \frac{\Delta P}{P} = \frac{\Delta Q}{Q} \times \frac{P}{\Delta P} = \frac{P}{Q} \times \frac{\Delta Q}{\Delta P}$$

so,

$$\frac{P}{Q} \times \frac{\Delta P}{\Delta Q} = 1$$

and

$$E_s = 1$$

A supply curve which passes through the vertical (price) axis is elastic, and one which passes (or, if extended, would pass) through the horizontal (quantity) axis is inelastic. This holds regardless of the slope of the curve, and it applies to the whole curve when this is linear (i.e. forms a straight line).

These statements can be proved by the same method as in Figure 5.13. Do not worry if you cannot prove them yourself – just remember the position. Examples are given in Figures 5.14 and 5.15.

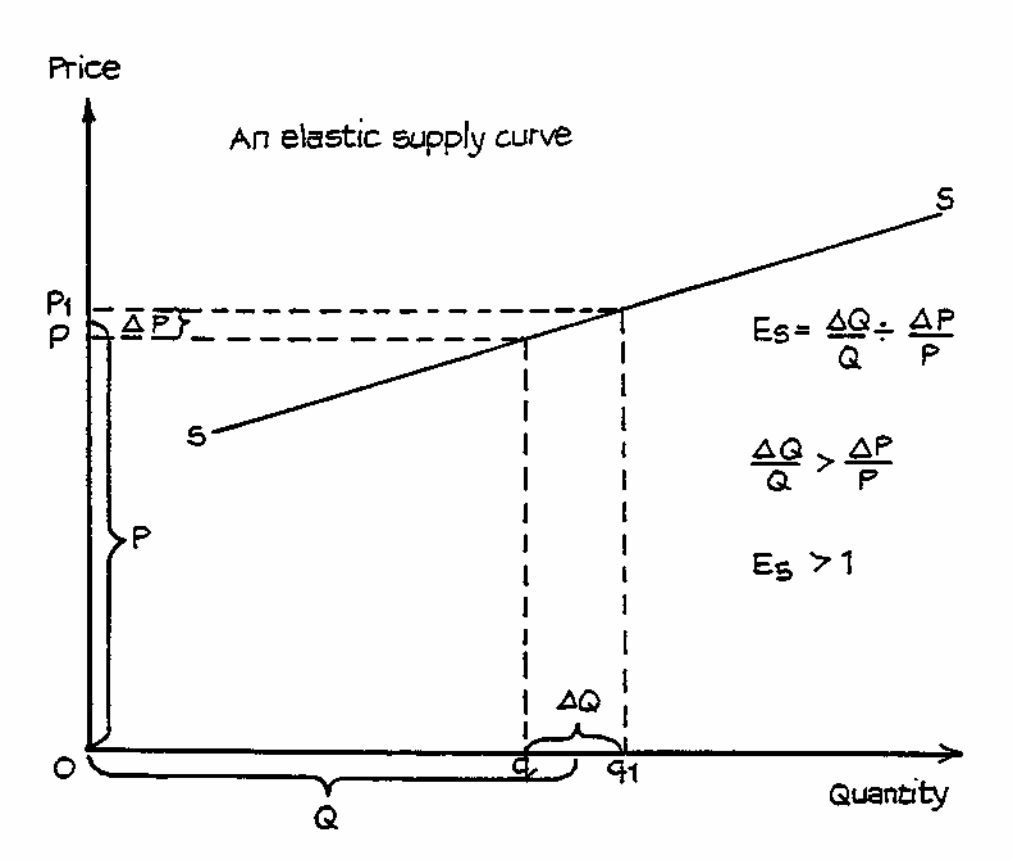


Figure 5.14: An elastic supply curve

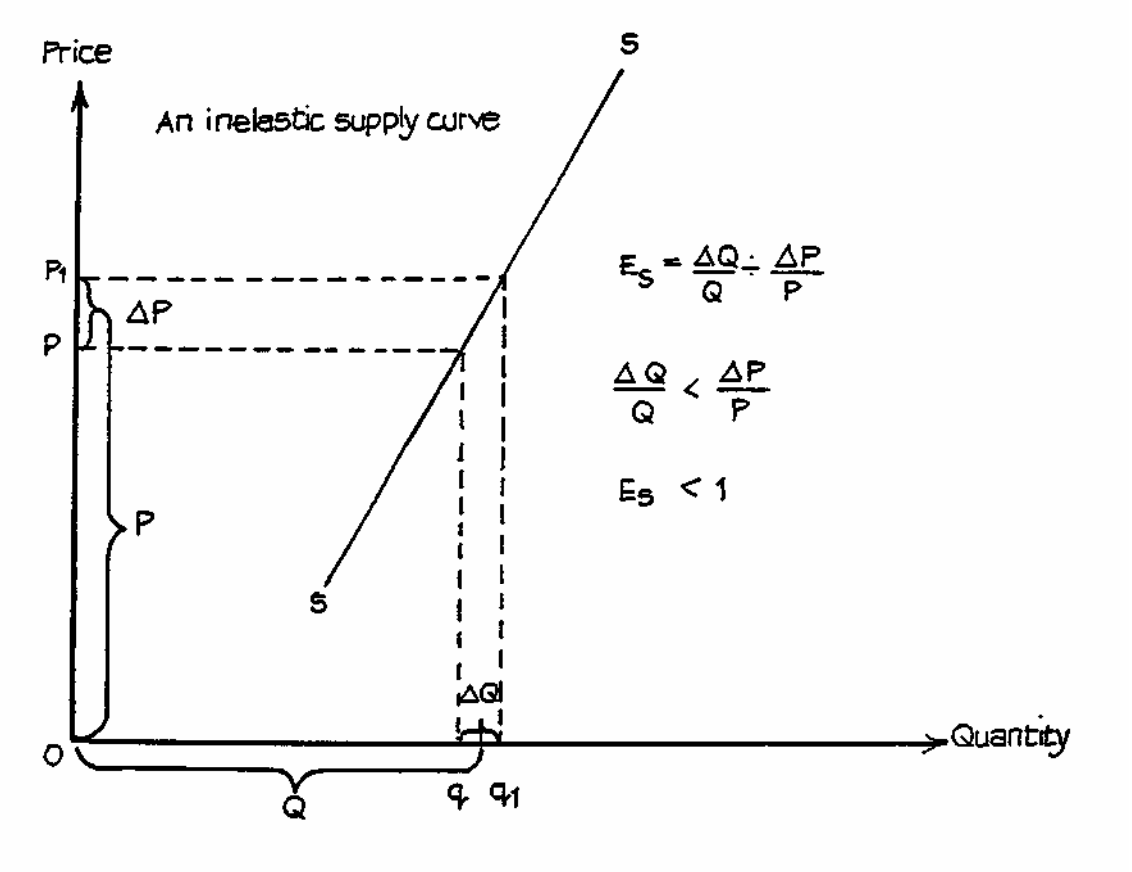


Figure 5.15: An inelastic supply curve

When the curve is non-linear, the important point is the direction of the tangent to the curve at the price level under consideration. This is shown in Figure 5.16.

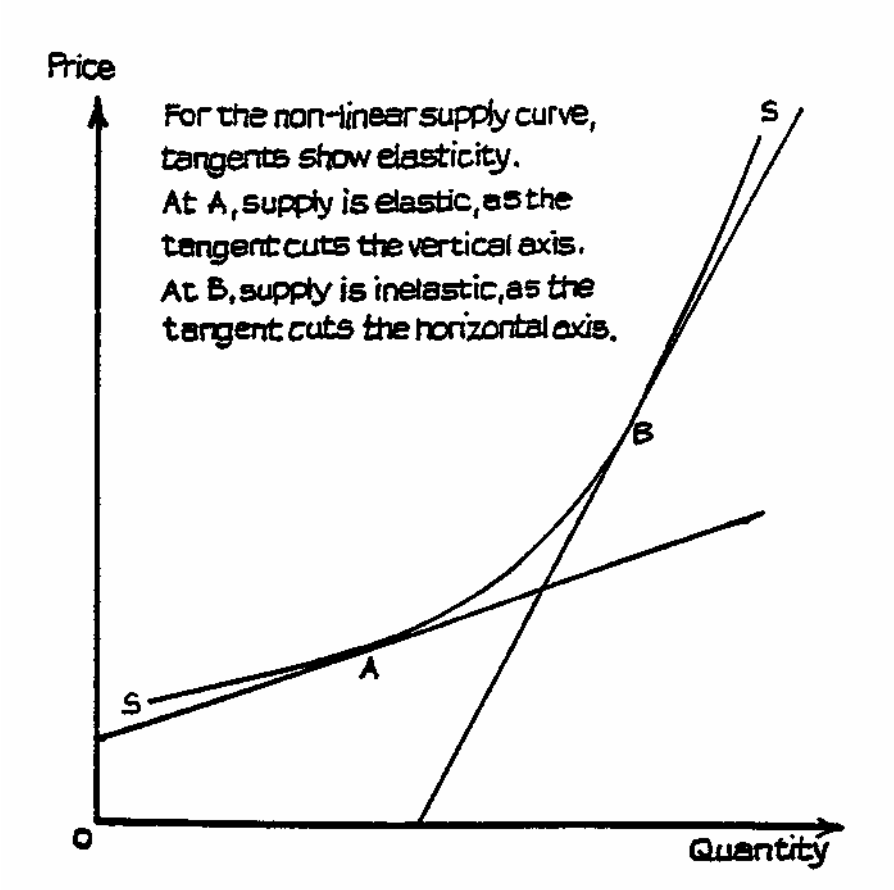


Figure 5.16: A non-linear supply curve

Elasticity of Supply in the Long Run

The main influence on the elasticity of supply is the speed with which producers can respond to changes in cost, price and profitability. Few firms can alter their production plans immediately when basic materials, capital and labour have already been committed to them. However as time goes on plans can be changed, workers can be hired or fired, and new machines bought or old ones scrapped.

The speed and ease with which production plans can be changed depends on the nature of the production process. As a general rule processes (such as services) which are labour-intensive can be changed more quickly than those that are capital-intensive. Workers, especially if they are part-time, can have their working hours increased or reduced and the number of workers employed can be changed; whereas capital-intensive processes, such as motor-vehicle assembly lines, still have to pay costs of capital even when equipment is no longer used. It may therefore be better to maintain production as long as variable costs are covered by sales revenue and there is some contribution to unavoidable fixed costs, rather than suffer the heavy losses of a major production change. However when the decision has to be made to reduce production the consequences can be swift and far-reaching, with large numbers of workers suffering redundancy.

We can say then, that supply will be inelastic in the short run and elastic in the long run. What constitutes short run and long run depends on production methods. Nevertheless, supply is unlikely to be completely inelastic even in the very short term, as some adjustment is usually possible. Even the motor-assembly track can be speeded up or slowed down in a matter of hours, in response to a managerial decision.

The change in elasticity over time is illustrated in Figure 5.17.

© ABE and RRC

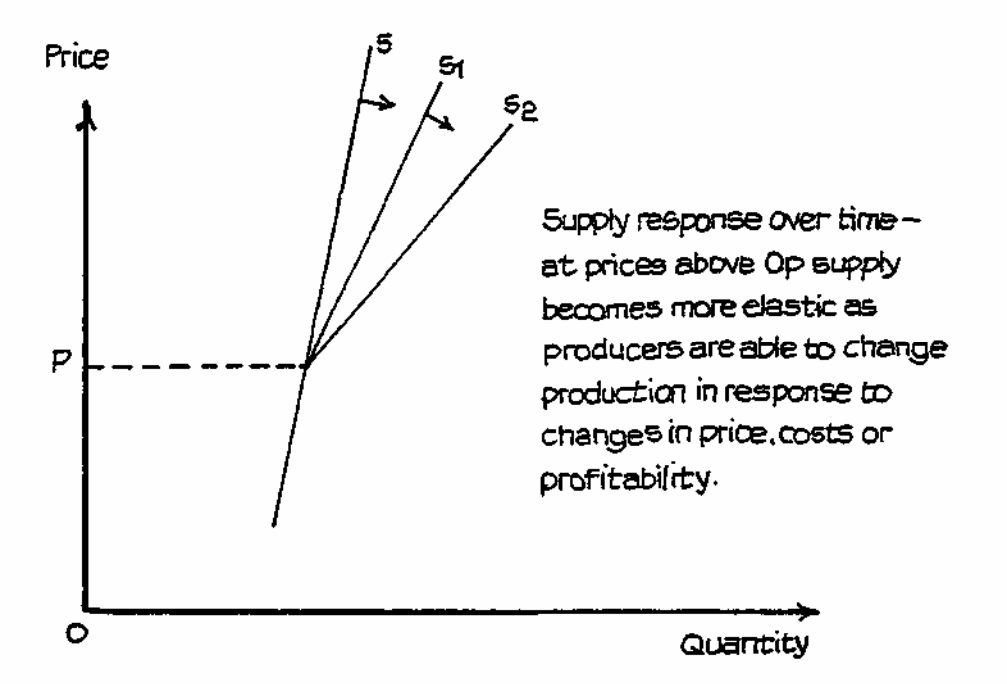


Figure 5.17: Change in elasticity over time

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. Which is the simplest definition of profit?
   (i) The rate of interest paid to savers.
   (ii) The excess of revenue over cost.
2. To maximise profits which output level should the firm produce at?
   (i) Average cost is equal to average revenue.
   (ii) Marginal cost is equal to marginal revenue.
   (iii) Total cost is equal to total revenue.
   (iv) Marginal cost is equal to average cost.
3. Which does the following formula calculate:
   (i) the elasticity of demand for a good or
   (ii) the elasticity of supply of a good?

$$\frac{\text{proportional change in quantity supplied}}{\text{proportional change in the product's price}} = \frac{\Delta Q_s}{Q_s} \div \frac{\Delta P}{P} = \frac{P \Delta Q_s}{Q_s \Delta P}$$

4. Does elasticity of supply measure the responsiveness of a firm's supply to changes in:
   (i) the market price of its product or
   (ii) its rate of profit?
5. Is the main influence on the elasticity of supply the speed with which producers can respond to changes in:
   (i) the slope of their supply curve or
   (ii) cost, price and profitability?
6. Is the general shape of a firm's supply curve:
   (i) downward sloping or
   (ii) upward sloping?
7. Firms will supply more output if they think it will lead to an increase in their:
   (i) sales
   (ii) profit
   (iii) elasticity of supply
   (iv) elasticity of demand?
8. A firm should cease production in the short run if its selling price does not enable it to cover all its:
   (i) average fixed costs
   (ii) average variable costs?

© ABE and RRC

© ABE and RRC

Study Unit 6

Markets and Prices

Contents ***Page***

A. Nature of Markets **103**
The Economic Good 103
Market Area 104
Communications and Transport 104
Conditions of Supply and Demand 104

B. Functions of Markets **105**
Information 105
Establishing Price 105

C. Prices in Unregulated Markets **106**
Definition of Unregulated Markets 106
Equilibrium Price 106
Changes in Intentions – Shifts in the Curves 107

D. Price Regulation **110**
Reasons 110
Effects of Price Controls 110

E. Defects in Market Allocation **112**
External Costs and Benefits 112
Public Goods 114
Inequalities of Income 115
Market Power of some Large Suppliers 115
Deficiencies in the Supply of Public Goods 115

F. The Case for a Public Sector **116**
Education 116
Health Care 116

(Continued over)

© ABE and RRC

G. Methods of Market Intervention: Indirect Taxes, Subsidies and Market Equilibrium **117**
What are Indirect Taxes and Subsidies? 117
Effect on Supply 118
Effect of Tax on Price 119
Subsidies 120
Government Use of Indirect Taxes 121

H. Using Indirect Taxes and Subsidies to Correct Market Defects **122**

© ABE and RRC

Objectives

The aim of this unit is to: explain the concept of market equilibrium and examine, using demand and supply analysis, the effects of changes in economic factors upon equilibrium price and quantity; explain the difference between private and social costs, and examine the consequences of externalities for the market equilibrium; examine the effects of various types of government intervention on market outcomes.

When you have completed this study unit you will be able to:

- explain, in words and diagrams, the concept of equilibrium in a supply and demand model, and the process by which equilibrium is reached
- examine the effects of changes in market conditions (for example a change in the price of a substitute good, a change in consumer income, an increase in advertising expenditure, the introduction of new cost-reducing technology) which lead to shifts in the demand and/or the supply curve upon the equilibrium; explain the importance of elasticity to the impact of such changes
- draw supply and demand curves based on data and solve for the equilibrium price and quantity
- explain the meaning of positive and negative externalities, and the distinction between private and social costs and benefits
- identify real world examples of externalities and discuss how they arise
- demonstrate the effects of externalities on the market equilibrium using demand and supply analysis and identify the social costs associated with the distortions caused by externalities
- demonstrate how taxation policy can be used to remedy problems caused by externalities and discuss the merits of a tax approach relative to possible alternative policies
- examine, using appropriate diagrams, the effects of taxes and subsidies on the market equilibrium, identifying the burden/benefits of taxation/subsidies on consumers and producers
- examine, using appropriate diagrams, the effects of quotas, price ceilings and price floors on the market price and quantity traded.

A. NATURE OF MARKETS

In economics, a market is an area within which the forces of demand and supply for a particular "economic good" can communicate and interact, so that the "good" can be transferred from suppliers to buyers.

This definition contains a number of important elements which have to be considered whenever we analyse a particular market or compare one market with another. Let us look at these elements.

The Economic Good

A good is any benefit which accords utility to people, and to obtain which they are prepared to sacrifice scarce resources. The term "utility" is chosen because it avoids the idea that there has to be any particular virtue in the good. If people want something and are prepared to make some sacrifice of their resources (usually represented by money) to obtain it, then we assume they gain utility from it, even if it does them actual harm. Thus economists may analyse the markets for tobacco or heroin.

© ABE and RRC

The good can be a physical object, such as a motor car, or it can be a service. It can be a consumer good, an intermediate good, a capital good, or a factor of production. In this course we are concerned chiefly with consumer and production factor markets.

We must be careful to give a precise definition of any market we are considering. The total market for motor cars contains a number of subsidiary markets – e.g. for sports cars or saloon cars. We must always distinguish the market for the whole class of product from that for a particular brand or other subdivision. Thus, the market for the Mini Metro is distinct from the market for small cars – which, in turn, is distinct from that for private cars and from the market for personal transport as a whole.

Confusion sometimes arises when we are concerned with the price elasticity of demand for a product. The class or product may be price inelastic, whereas a particular brand may be price elastic. For example, petrol in general may be price inelastic, but the price of K's petrol can be price elastic. The motorist has to have petrol, but she may have the choice of a number of filling stations offering a variety of petrol brands at different prices, and she may also be prepared to go a few miles out of her way to obtain the cheapest brand of petrol.

Market Area

We need to examine the market area when considering the conditions of a particular market. The area is that within which communication takes place, and not simply where final negotiation is arranged. A sale of antiques or fine paintings may take place in a small room in London. However beforehand catalogues may have been sent to dealers throughout the world, and many foreign buyers may be represented by their agents when the sale or auction actually takes place. In contrast, a small retail shop may be concerned with a market area restricted to a few streets or a single housing estate. The goods it sells may be available in other shops serving different market areas nearby.

Communications and Transport

The extent of the market is really determined by the efficiency of communications and the ability to transport the goods from seller to buyer. X does not really have a choice between goods A and B if he does not know that B exists, or if he has no means of comparing price or quality. Thus, if I am buying tomatoes on one side of the town, I cannot really compare them with those on sale on the other side of the town, even if someone tells me that they are several pence cheaper. I need to be sure that they are products of similar quality.

Some markets have developed very precise descriptive terms. The use of these terms, for example in some of the basic commodity exchanges, enables buyers and sellers to know exactly what quality goods are being traded.

There can be an effective market only if it is possible to transfer the product from seller to buyer. Any barrier to transfer will limit the market area.

Conditions of Supply and Demand

There can be a market only if there are suppliers able to deliver the goods at the time agreed, and buyers with the necessary resources to acquire them.

The good does not necessarily have to be in existence at the time it is traded, as long as there is a guarantee that it will be available when and where agreed. The ability of certain commodity markets to trade in crops not yet grown, or metals not yet mined, is well known; but a manufacturer can also agree to sell goods not yet made, and a few authors can even sell books not yet written! However, both buyer and seller must have a clear idea of the product that is to be delivered. The more precise the definition of a product, the easier it is to sell in this way.

© ABE and RRC

The desire to buy must also be realistic. Many of us would like to possess an ocean-going cruiser or a private aeroplane; but few of us have the resources to acquire and operate them.

B. FUNCTIONS OF MARKETS

A market has other purposes, apart from providing the means whereby a good is transferred from supplier to buyer.

Information

The market serves to convey information about the conditions of supply and demand. I may go to a furniture store, not just to buy a piece of furniture but to see what furniture is available and at what price. The better the communication system within the market, the more information I can gain about what can be bought – and the more chance I have of achieving full utility from my purchase.

This communication function works both ways. The market also informs actual and potential suppliers about the strength and pattern of demand – about what people want to acquire and what level of price they are prepared to pay. Suppliers need this information in order to plan production.

The problem from the supplier's point of view is often that the information comes too late. The supplier has to make supply decisions before accurate information is available. The supplier wants to know today what market conditions are going to be like tomorrow. The impossibility of achieving accurate forecasts all the time is one of the main sources of business risk.

Establishing Price

Arising out of the two-way communication function is a further most important function – that of establishing the price at which the buyer is willing to buy and the supplier willing to supply. How this may be achieved is the subject of much of the rest of this study unit.

It is such an important function of the market that some large firms ensure that certain markets continue to operate only because they need a reliable mechanism for price-setting. The large manufacturing companies do not really need to buy metal on the London Metal Exchange – they can obtain all they need direct from suppliers. But they do need to know the conditions of demand and supply in the main areas where metal is bought and sold. By keeping the metal exchange in operation, they obtain this information, which provides a price-setting mechanism and so helps to reduce some of the uncertainties which they have to face in obtaining essential materials.

© ABE and RRC

C. PRICES IN UNREGULATED MARKETS

Definition of Unregulated Markets

The term "unregulated" here means not subject to any price-setting regulation. An unregulated market can be subject to detailed regulations regarding the conditions of payment and transfer and the procedures for settling disputes. However these assist rather than impede the free communication of buying and supplying intentions, and allow them to interact in order to establish a market price. An unregulated market is thus one in which the forces of supply and demand are free to interact, without any form of outside price control.

We tend to think of regulation in terms of control by the State or its agencies, but of course a market can be controlled in other ways. Certain local antiques auctions are reputed to have been controlled by rings of dealers who agree not to bid against each other and to share purchases among themselves after the auction. This is not an unregulated market! The prices paid for goods at such an auction are not "market" prices because they do not reflect the true conditions of demand.

Equilibrium Price

The equilibrium price is the one at which the intentions of suppliers are just matched by the intentions of buyers, i.e. where the amount of the good demanded is just equal to the amount provided. In this state there is no pressure from either supply or demand to move away from this price, so the market forces are in a state of rest – in equilibrium.

We have examined the concepts of supply and demand schedules and curves. If we put supply and demand schedules and curves together, we can arrive at the equilibrium price, i.e. *the market price*.

Suppose we have the supply and demand schedules for the (fictitious) product Whizzo, as set out in Table 6.1 and illustrated in Figure 6.1.

Price per kilo	**Quantity (kilos per week)**	
£	**Producers willing to supply**	**Consumers willing to buy**
1.50	200	700
2.00	300	675
2.50	400	650
3.00	500	625
3.50	600	600
4.00	700	575
4.50	800	550
5.00	900	525

Table 6.1: Supply and demand schedules for Whizzo

© ABE and RRC

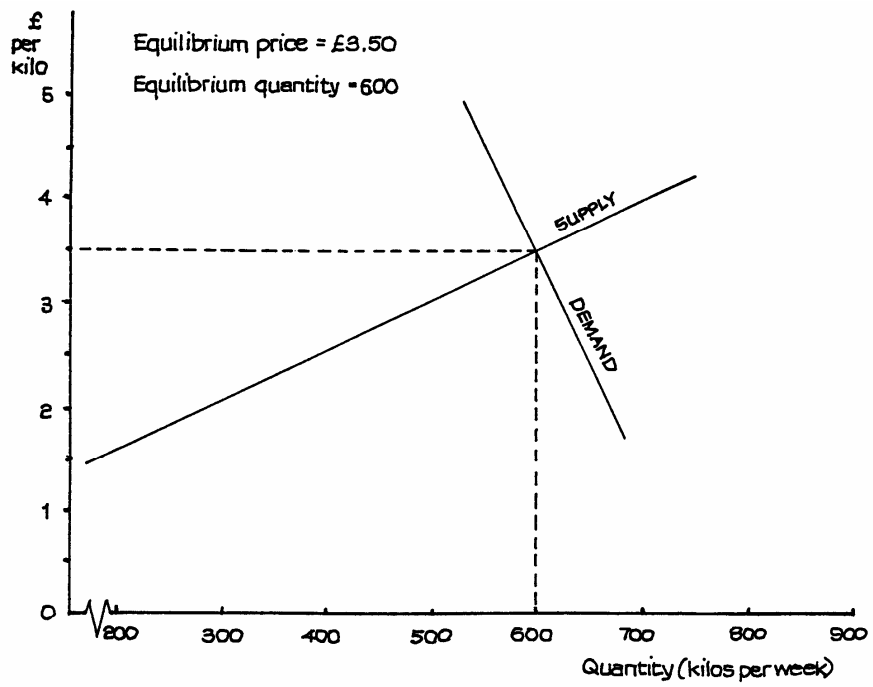


Figure 6.1: Supply and demand for Whizzo

We can see from the schedules and the graph that it is only at price £3.50 (600 kilos per week) that the intentions of producers and buyers are the same. At any higher price, producers will be supplying more than buyers are willing to buy. At any lower price, producers will not be supplying enough Whizzo to meet demand. The equilibrium price is £3.50, and 600 kilos per week the equilibrium quantity. As long as neither set of intentions changes, there is no incentive for any movement away from this price and quantity, once it is achieved.

Changes in Intentions – Shifts in the Curves

We can show the concept of equilibrium price and quantity in a general graphical model, as in Figure 6.2. Here, equilibrium price is Op and equilibrium quantity Oq – the price and quantity level where the supply and demand curves intersect. We can develop this approach to analyse the result of movements in the supply and demand curves.

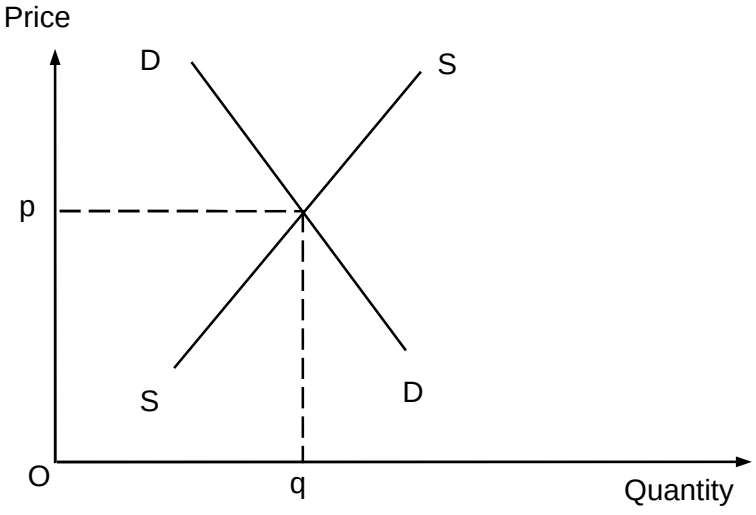


Figure 6.2: Equilibrium price and quantity

(a) Change in Either Demand or Supply

Look at Figure 6.3. Here there is a shift in buyers' intentions, caused perhaps by a change in taste, supported by an increase in advertising. The result is a movement of the demand curve from DD to D_1D_1.

In this model, supply intentions remain unchanged. The result is an increase in the equilibrium price and quantity from Op, Oq to Op_1, Oq_1.

We can use the same technique to illustrate the effect of a shift in suppliers' intentions. This is shown in Figure 6.4, where supply falls from SS to S_1S_1. Demand intentions remain unchanged (DD) and the equilibrium price and quantity move from Op, Oq to Op_1, Oq_1.

Price rises and quantity traded in this market falls.

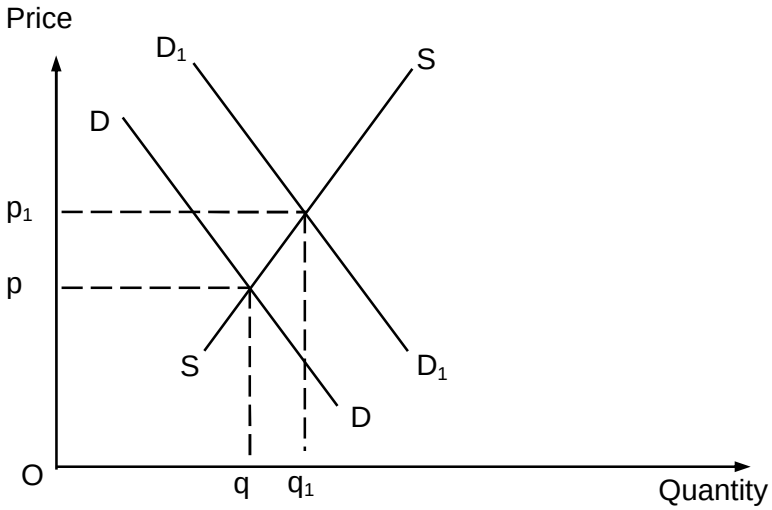


Figure 6.3: Movement of the demand curve

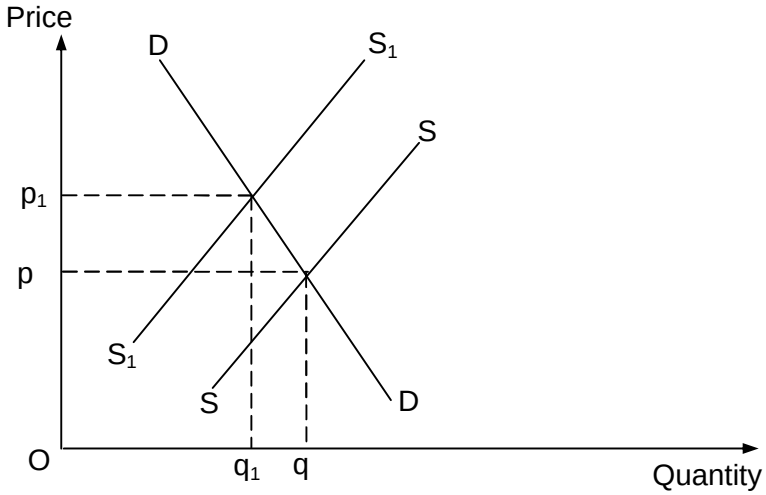


Figure 6.4: Movement of the supply curve

(b) Change in Both Demand and Supply

So far we have considered only a possible shift in demand or supply. In practice, a movement in one is likely to influence the other through the effect on price and quantity.

Suppose there is a major increase in demand, represented by a movement of the demand curve in Figure 6.5, from DD to D_1D_1. This shift, if supply remains unchanged at SS, results in an increase in equilibrium price from Op to Op_1, and in quantity from Oq to Oq_1.

Now suppose that this increase in quantity makes it worthwhile for one or more producers to develop new production methods, so that the good can be mass-produced at a lower unit cost. The result, after a time interval, is to shift the supply curve from SS to $S_{t+1}S_{t+1}$. Here the $t + 1$ indicates a change in time period.

The new supply schedule, combined with the increased demand, produces a fresh equilibrium price and quantity at Op_{t+1}, Oq_{t+1}. We have the apparently unusual result of an increase in demand resulting in a reduction in market price. Note however that this can happen only when given some rather special assumptions about the stage of a product's development and the possibility for change in supply conditions.

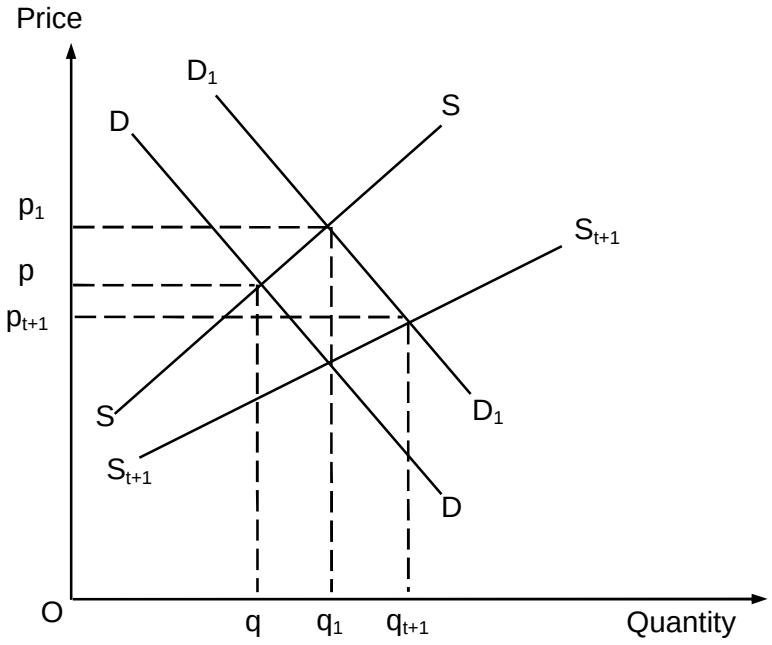


Figure 6.5: Movement of both the demand and supply curves

Normally, we expect an increase in demand to raise equilibrium price and quantity. This is the direct effect. The *later* reduction in price can result only from a shift in the supply curve, indicating a completely new set of supply conditions.

A somewhat similar process can be initiated by a change in technology, allowing mass-production at a reduced price. Here, there is first a shift outwards in the supply curve. Demand then rises but not enough to stop the price from falling. Consider the market for mobile phones in this light.

© ABE and RRC

D. PRICE REGULATION

Price regulation refers to the imposition of a minimum or a maximum market price by government decree or international agreements/organisations, such as OPEC. A maximum price is set by the imposition of a *price ceiling*. A minimum price is set by the imposition of a *price floor*.

Important applications of such price ceilings and floors include minimum wage legislation, maximum prices for some food items and/or fuel, maximum prices for rented accommodation, and minimum and maximum prices for some commodities in international markets.

Reasons

If price and quantity will always move to equilibrium provided economic markets are left alone, we must ask why governments and other agencies should ever wish to intervene. In practice, there are several reasons, of which the following are among the most common.

(a) Social Unacceptability

If the price resulting from an unregulated market were considered to be socially unacceptable, as causing hardship or conflict in the community, attempts might be made to control it. This could happen in a period of food shortage caused by war and/or climatic disaster, and also if there were a shortage of housing in urban areas sufficient to cause hardship and increase risks of disease, crime and other social evils.

(b) Incomes of Producers

Attempts might be made to maintain high prices if it were desired to raise the income of producers and their employees. This is one of the motives of the European Union's Common Agricultural Policy (CAP).

(c) Stability of Supply

Some markets are notoriously unstable because of unplanned variations in supply, caused by weather and other circumstances beyond the control of producers. In these cases, attempts may be made to control prices to ensure greater stability in the market.

Effects of Price Controls

If prices are controlled without any attempt to control demand and/or supply at the same time, the result can be the opposite of that intended. This is illustrated in Figure 6.6.

© ABE and RRC

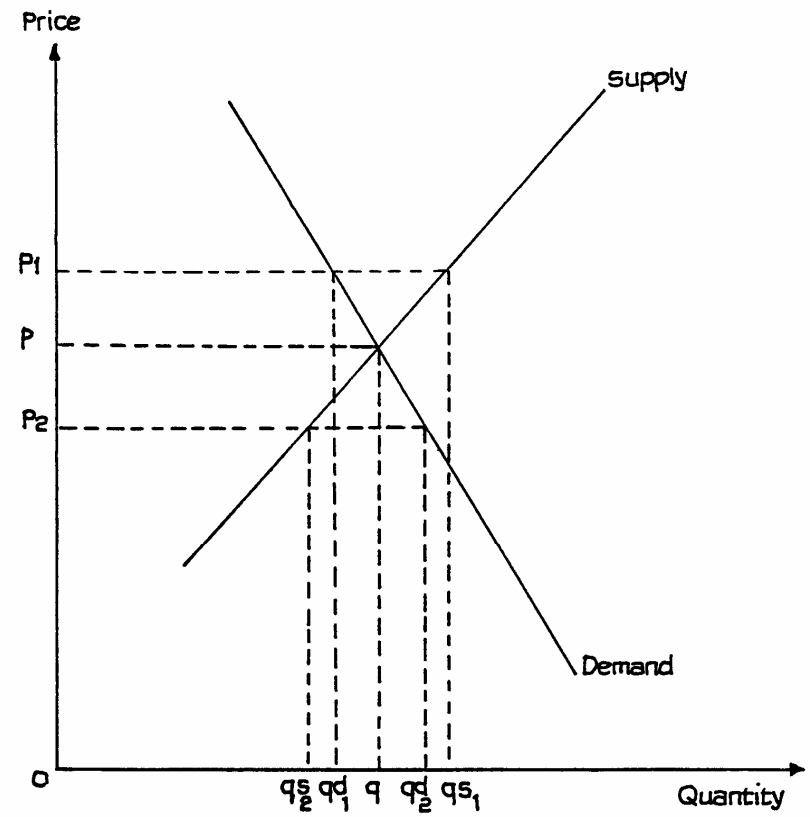


Figure 6.6: Supply surplus and shortage

Looking at the diagram, if price is fixed at p_1, quantity supplied (qs_1) is more than that demanded (qd_1), and there is surplus production.

If price is fixed at p_2, quantity demanded (qd_2) is more than that supplied (qs_2), and there is a shortage.

Only at price p will quantity supplied equal quantity demanded.

Here, we see that any attempt to fix prices at a level other than the market equilibrium price of p will produce either surplus production (fixed price $p_1 > p$) or a shortage (fixed price $p_2 < p$).

We are forced to the conclusion that on their own, price controls are ineffective. Governments and other bodies must identify the real problem and seek to solve that. For instance, if the problem is lack of adequate supply (say food or housing shortage), then the government must either increase supply, e.g. by making additional payments (subsidies) to suppliers, or by entering the market as a producer or importer. If these remedies are impossible, the government must ration the available supply among consumers in a way that the community regards as acceptable.

Such measures may be effective, at least for a time, though they may be expensive to administer and police. The government or other agency must decide whether the social benefits to be gained from market regulation justify the cost and opportunity costs of the resources used in maintaining the regulations. Care must also be taken to ensure that the regulations themselves do not discourage suppliers to the extent that the basic objects of the policies are defeated. The heavy bureaucracy created by many schemes in the so-called planned or socialist economies often significantly discourages total production. If the problem is excess supply, then the government may seek either to stimulate demand (e.g. by reducing prices through the payment of subsidies), or to reduce supply by encouraging or

© ABE and RRC

paying producers to leave the market (as in the case of European Union measures to reduce European milk and wine supplies).

The most difficult problems often involve unplanned fluctuations of supply, when the plans of regulatory bodies can be upset by (say) unusually good or bad crops owing to weather conditions. If there are fairly regular cycles of overproduction or underproduction, and demand is reasonably constant, and if it is possible to store the crops, then the government can apply a mixture of controls over prices and production combined with purchases of overproduction to keep in store for release in periods of underproduction. However, it is found that the guaranteed prices that usually form part of such policies lead inevitably to steady increases in production. The government then finds itself storing quantities of goods that it has little hope of ever releasing for resale, except at very low prices to people in other parts of the world. It may even have to give away some of the surplus produce. Such policies then become a heavy burden on taxpayers and lead to hostility from the community.

It is clear that governments which embark on market-intervention policies may, and often do, find that they become involved in increasingly difficult and expensive measures that do very little to solve the problems they were meant to eliminate.

There are other reasons why governments may choose to intervene in the market to alter the resultant market equilibrium.

E. DEFECTS IN MARKET ALLOCATION

In very many cases, unregulated markets and the price system are effective and efficient ways of allocating resources. Also, as we saw in the previous section, some forms of well-meaning government intervention can actually make worthy social objectives more difficult to achieve. Nevertheless, this does not mean that unregulated markets are always perfect. The existence of some defects is widely accepted and we will now consider the main ones.

External Costs and Benefits

External costs and benefits are also referred to as "*externalities*". Externalities or external effects are very important because they give rise to *"merit goods"*, *"demerit goods*" and *"public goods*".

External Costs

Not all the costs of factors used in the production process are paid by the producer as private costs. For example, suppose that during a dry summer, a farmer watered his crops with water pumped from a canal. As a result, the canal level fell and it could no longer be used by waterway travellers. Unless the farmer paid compensation to the travellers, it is clear that they would be contributing to the costs of the farmer's production. Because these costs are being paid by people external to the production process, they are called "external costs".

We can think of many examples of such costs, for instance road users who incur additional fuel and machine-wear costs resulting from motorway delays. If these delays are caused by repairs needed to make good damage brought about by heavy lorries travelling at high speeds, then other road users are contributing to the costs of transporting goods by these lorries. If a proportion of the cost of road repairs is paid from general taxation, then all taxpayers are contributing to the costs of road travel – even those tax payers who rarely travel at all.

Other examples of external costs include the poisoning of rivers by industrial waste, the pollution of sea coasts by waste oil discharged by oil tankers, the sickness and early deaths of workers from industrial diseases. The list is almost endless, and you can probably add to it from your own observation. Some costs may even be borne by later generations. The most

© ABE and RRC

serious example of an external cost confronting the world today is that of global warming, caused by atmospheric pollution from the continued and excessive burning of oil and coal.

The existence of an external cost associated with the consumption of a good such as alcohol or cigarettes means that the social benefit is less than the private benefit from consumption. Such goods are examples of *demerit goods*. Because consumers ignore the *negative externalities or social costs* created by their consumption of such goods, they are overproduced and over-consumed in a free market without government intervention.

External Benefits

In contrast, it is possible for people to receive benefits from production towards the cost of which they have not contributed. These are external benefits. If a large firm builds modern roads or provides other transport facilities which are then available for use by the general community, then that community gains external benefits. If a business firm provides a good canteen and housing for its workers and, by improving standards of housing and welfare, improves the health of workers and their families, then this, too, is an external benefit. We are well aware of cases where firms cause damage to the environment, but there are also cases were firms improve the environment by renovating property, creating sports grounds, or even parks.

The existence of an external benefit associated with the consumption of goods/services such as health care and education means that the social benefit is greater than the private benefit from consumption. Such goods are examples of *merit goods*. Because consumers ignore the *positive externalities or social benefits* created by their consumption of such goods, they are underproduced and under-consumed in a free market without government intervention.

Economics of Externalities

It might be thought that economists would favour external benefits and dislike external costs. In fact, economic theory suggests that all externalities distort the use of resources, and that even external benefits are probably better provided in other ways. The danger of external costs can easily be recognised. For example, if road users, especially heavy goods vehicle users, do not pay the full costs of their road use but pass some of these on to the rest of the community, then the relative costs of transporting goods by road – as opposed to by rail or water – are distorted in favour of road. Consequently, goods are carried by road transport at a higher cost to the community than it would have paid if they had been carried by other means, say by rail. The community is not making the most efficient possible use of its available resources, and its living standards are lower than they would otherwise be because some production is being lost. Moreover, in situations of this type, the problem tends to be self-worsening. If road transport is artificially cheap, then goods are diverted to road from rail. Road services are overcrowded, and there is pressure to devote more land to roads. Rail services are underused. Agricultural and residential land is lost to roads to carry traffic which could otherwise have been carried by substitute services.

This is what we mean when we say that externalities distort the use of scarce economic resources.

Externalities and the Government

What can be done about externalities? Does the community just have to accept their existence? Clearly neither the producers who are able to pass costs to others, nor the buyers of their goods or services who obtain reduced prices because of the reduction in private costs, are likely to volunteer to pay more unless they are obliged to do so. They could not do so as individuals in competitive markets. Only governments, acting on behalf of the community as a whole and reacting to political pressures, can take effective measures. The options open to government are the following:

- Legislate to make actions considered undesirable illegal, and enforce the law. In a democracy such laws must be acceptable to the community as a whole; care must be

© ABE and RRC

taken to ensure that desirable benefits are not lost and that the cost of law enforcement is not out of proportion to the costs avoided.

- Legislate to ensure that producers behave in a socially acceptable way and follow practices designed to avoid the undesirable external costs. Water and sewerage companies may be required to achieve certain minimum standards. The costs of complying with the law thus become private costs and part of the production cost which must be met by users of the goods and services. All producers then become subject to the same requirements so that none can gain a competitive advantage by not complying with the standards. If producers have to compete with foreign imports the government will have to ensure that these imports are subject to the same minimum standards.
- Impose special taxes designed to make some products very expensive and so discourage their use. There are several objections to this course of action. The government might start to rely on the revenue from the taxes and so take care to keep them at a level where the products are still bought and used; the taxes may well then cease to deter or reduce the external costs. Alternatively the government might impose very high taxes with the result that there is widespread tax evasion; the cost of collecting the tax and punishing evaders then rises to impose additional burdens on the community.
- Pay subsidies to suppliers to reduce the market price paid by consumers and thereby encourage increased consumption of merit goods. Alternatively, the State may take overproduction and ensure, through legislation, that all the relevant consumers are provided with the socially optimal level of the good or service. For example, compulsory school education is provided by governments in many countries.
- Clearly it is more desirable to try and ensure that external costs are removed altogether rather than that they should simply become private costs. Even if employers are forced to pay adequate compensation to workers whose lungs are damaged by dusty manufacturing processes, the workers still suffer. However, if manufacturers are required to have efficient dust extraction equipment, private costs are increased but the health of the workers is improved. At the same time care must be taken to ensure that external costs are not simply exported. For example, one way of dealing with dangerous gases might be to ensure that they are expelled through very high chimneys, but unfortunately these may simply redirect the gases to another country for that country to bear the cost.

There is no universal and simple method of dealing with externalities. On the whole it does appear that the market economies have been more successful in controlling and reducing undesirable external costs associated with environmental pollution than have the old command economies. This is probably because in the more open and consumer-orientated societies, producers and government have had to be willing to respond to pressures from the public when that public has been determined to eliminate socially unacceptable practices.

Public Goods

Merit and demerit goods are produced in a free market, without government intervention; the problem is that either too little or too much is produced. Too few merit goods are consumed in a free market because consumers ignore the external benefits associated with such goods. In contrast, there is over-consumption of demerit goods in a free market because consumers ignore the external costs. In the case of "*public goods*" the market failure is that the goods are not produced at all if left to the free market.

Most goods and services, including merit and demerit goods, are private goods and services in the sense that if they are consumed by one person their availability is correspondingly reduced, and one person's consumption cannot be consumed by another person. Public

© ABE and RRC

goods are different. Pure public goods are defined as those goods or services which have the characteristic that one person's consumption does not reduce the amount available for consumption by others.

The alternative, and more revealing, way of looking at this characteristic is to note that if such a good or service is provided for just one person the supply is also freely available for consumption by others! What this means is that whoever pays for the production of the good or service is providing the same benefits for all others in society free of charge. The consequence of this is that no one is prepared to provide such a good or service because they are unable to recoup some of the cost by charging others for their consumption of the benefits. Thus public goods are not provided in a free market without government intervention. Although there are very few if any examples of pure public goods, national defence and lighthouses are examples of goods that have many of the features of a public good.

Inequalities of Income

One of the virtues claimed for the unregulated market is that it makes the consumer sovereign and that resource allocation responds to demand pressures. However, if we imagine that consumers influence allocation by votes cast when they buy or refrain from buying goods and services, we have to admit that some consumers have more votes than others and large numbers have very few votes. Markets respond quickly to those groups which have the most purchasing power. This does not always ensure that resources are allocated in ways that meet the social expectations of the community.

It has always been difficult to ensure that the poorest sections of the community are adequately housed. Normal commercial suppliers of housing are unwilling to meet this demand because the people concerned cannot afford to pay the full "economic costs" of housing, i.e. it is not usually possible to make a profit from providing housing for the poor. It is much more profitable to provide second homes for the wealthy. Not only does this offend against many people's ideas of social justice, but the housing problem rebounds against the community. The community is faced with extra costs because inadequate housing leads to poor health, disease, crime and a wide range of social problems that become a charge on the taxpayers. Only the State can intervene to improve housing for the poor. It cannot do so simply by holding down rents. It has to promote supply either by setting up State suppliers or by subsidising private suppliers so that supply becomes profitable.

Market Power of some Large Suppliers

Consumers may not always be as powerful as introductory economic theory suggests. Later we will learn about markets dominated by large firms. If such firms become very powerful, they can influence both supply and demand through controlling the goods allowed into the market and by heavy advertising. Governments of most large market-economy nations are often accused of failing to take action to check the sale of tobacco and alcohol – both of which are potentially dangerous to health and society – because of the power of the tobacco and alcohol producing companies. Even more notorious is the extremely powerful gun lobby in the USA.

Deficiencies in the Supply of Public Goods

The market economy operates on the principle of self-interest. Consumers wish to maximise their own utility and producers their profit. In most cases this works to the public benefit but not always. If it is in no one's interest to provide a community or public good, it will not be provided without the intervention of the political machinery of the State. Public sewers, public roads and transport, police and social services, even fire services, fall into this class. The community clearly needs adequate services but left to the market only the wealthy would

© ABE and RRC

attempt to purchase their own, and the community as a whole would be subject to the risk of contagious diseases, unchecked crime and fires.

F. THE CASE FOR A PUBLIC SECTOR

In noting the defects of the market economy as a means of allocating resources we have, in effect, made a case for a public sector within which the State, through its political structures, makes good the gaps and deficiencies of the unregulated market. The State can ensure that there is a minimum standard of housing for those with low incomes, build roads and establish communication systems. It can build sewerage systems and a system for piped, clean water, and provide police and fire services. It can provide a health and education service to ensure that all who are sick obtain medical care regardless of income and all children achieve a minimum level of education essential for survival in the modern world.

In communities with high living standards the question then arises as to *how far State provision should go* in the provision of public goods which at some stage tend to become private goods.

Let us take a closer look at two particular, high-profile issues.

Education

Most would accept the need for all to receive a basic education, but this does not necessarily mean that all who wish to do so should have the right to free education to doctorate level. Since there is evidence that, on average (but not, of course, for all individuals) there is a correlation between income level and length of time spent in full-time education, then education beyond the minimum represents a personal capital investment; many would argue that such education should be paid for by those who will benefit from it. Counter arguments are that the community benefits from the contribution of its most highly skilled and educated members (e.g. brain surgeons). The community should therefore pay to obtain the maximum potential from its scarce human resources; also those who earn high incomes normally pay the most taxes and thus pay eventually for the education they have received. There is no clear right or wrong answer to this debate, but you can see that the precise boundaries between the public and private sector in the supply of goods such as education are not clear-cut and the matter is arguable.

Health Care

Another area of public controversy is the provision of health care. The community clearly needs a health service, if only to defend itself against dangerous diseases which could quickly become plagues if large numbers of people could not afford treatment. Most people's ideas of social justice would accept that a person stricken by accident or sickness should receive treatment regardless of income. However, should this mean that all forms of treatment should be available for all regardless of income? Should the diseases of greed and overindulgence be given the same care as those of poverty and ignorance? If people can afford to pay for additional treatment or for more comfortable treatment, or non-urgent treatment at times that suit them rather than at times that suit a bureaucratic administration, is there any reason why they should not do so? No one passes moral judgment on those who choose to spend their income on exotic holidays rather than a fortnight at Benidorm, yet many pass such judgment on those who prefer to pay for a private room when they are in hospital instead of sharing a public ward.

Clearly many of the arguments surrounding health care involve emotionally charged value judgments resulting from past social injustices and history, but there are also serious economic considerations involved. The economist is concerned with the allocation of scarce resources, and we have to recognise that resources devoted to health care are scarce. The

© ABE and RRC

march of technology and medical science has made possible cures and treatments unimaginable when the National Health Service commenced in the 1940s. Open heart and transplant surgery require a massive investment in resources but benefit only a relatively few people. The proportion of old people is far greater than in the 1940s and the demands they make for health care are proportionally much greater also. Not even the most wealthy and advanced nation can provide all the resources that would be required to give immediate treatment to all those wanting it. Difficult allocation decisions have to be made and are made daily.

It can be argued that a private health system which permits scarce resources to be allocated on the basis of ability to pay, or by virtue of employment in a company that provides health insurance as part of its remuneration, is diverting resources from areas of greater personal or social need. One person suffers pain so that a consultant can earn a private income treating a less urgent patient in a private hospital. On the other hand it can be argued that the private health service brings in resources that would otherwise not be available. The consultant is willing to work for a relatively low level of pay from the National Health Service because he or she can have the additional income from private patients. Without this, the best surgeons would possibly go to countries where earnings were higher. Private hospitals relieve the public health service of many patients and reduce its need for expensive capital equipment. The debate can again continue with no clear right or wrong.

The basic problem is really one of allocation of scarce resources: the public versus private health service is only part of a much larger economic and social issue which concerns to whom, how and on what basis resources should be allocated for health care. How should the community decide what proportion of available scarce resources should be devoted to the technically brilliant feats of surgery which bring acclaim to surgeons and enable them to attend conferences abroad, and how much to the unglamorous, humdrum work of caring for the mentally ill for whom there is no hope of cure and little chance of international laurels for the carer? The unregulated market will not provide an answer, nor will a medical service subject to all the usual human vanities and frailties. The answer must eventually come through the political machinery of the community and the quality of the answer will reflect the health of that machinery.

Similar issues can be applied to virtually every other public sector and public utility service, and you should give some thought to the allocation problems inherent in, say, police, fire, water, and housing services.

G. METHODS OF MARKET INTERVENTION: INDIRECT TAXES, SUBSIDIES AND MARKET EQUILIBRIUM

What are Indirect Taxes and Subsidies?

Governments often influence markets through taxes and subsidies.

- An *indirect tax* is one that is not levied directly on individuals or organisations but is applied at some stage in the production or distribution of goods or services. It therefore affects prices and so is paid indirectly, through price, by consumers and income-earners. For this reason indirect taxes are often referred to as expenditure taxes and are listed as such in the British national accounts which appear in the annual publication known as the *Blue Book of National Income and Expenditure*.
- *Direct taxes* are those levied directly on income or wealth as it is created and are paid by the income-earner or wealth-earner to the government. The economic implications of direct taxes are considered later in the course.

© ABE and RRC

At this stage it should be clear to you that anything that influences market price will have consequences for both supply and demand, with the result that the final consequences of a tax may not be what the government intended.

Sometimes, of course, a tax may be imposed with the deliberate intention of influencing supply or demand. More often it is levied as just another way to raise the revenue that governments imagine they need, and they seek to have as little effect as possible on the production system. In practice, any tax must have an impact, as we shall see.

A *subsidy* can be seen as a reverse or negative tax. It is a payment to a producer or distributor, so that its effect is to increase supply. So to judge the effects of a subsidy, simply reverse the arguments presented in relation to the tax – but remember of course, that in order to pay a subsidy, the government has to have revenue, and its main source of revenue is tax. Generally, then, a subsidy paid to A means that B and C have to be taxed. The harmful effects of the tax may outweigh any beneficial effect of the subsidy.

Effect on Supply

The effect on supply of an indirect tax being imposed is illustrated in Figure 6.7. This shows a supply curve SS, indicating that production can range from 200 units per week at a price of £4 to 800 units at a price of £10.

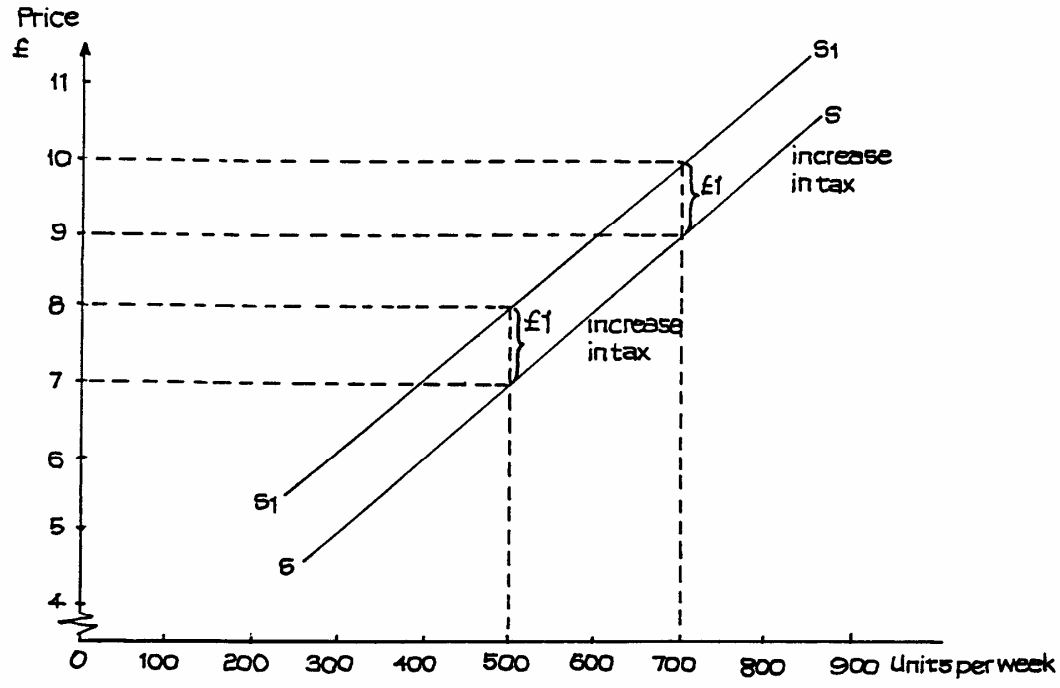


Figure 6.7: Effect of an indirect tax on supply

Suppose a new tax is imposed at £1 per unit. To supply 500 units per week, producers wanted a price of £7 per unit. After the imposition of the tax, the producers still want to receive £7, but to get this, the price has to rise to £8 to include the £1 per unit that now has to be paid to the government. Similarly, to keep production at 700 units per week, the price has to rise from £9 to £10 per unit.

Imposition of the tax thus moves the supply curve to the left (SS to S_1S_1). The vertical distance between the curves represents the amount of the tax.

Of course, a subsidy paid to the producer moves the supply curve to the right because the argument is exactly reversed.

© ABE and RRC

In Figure 6.7 the after-tax supply curve S_1S_1 is parallel to the before-tax curve of SS. This suggests that the tax or tax increase is flat rate, i.e. the same at all price levels. In practice indirect taxes such as VAT depend on value and are sometimes known as *ad valorem* taxes. Usually we would expect the tax to be expressed as a percentage of value or price, and its amount will therefore increase as price rises. In such cases the gap between the two supply curves will increase at the higher prices as illustrated in Figure 6.8.

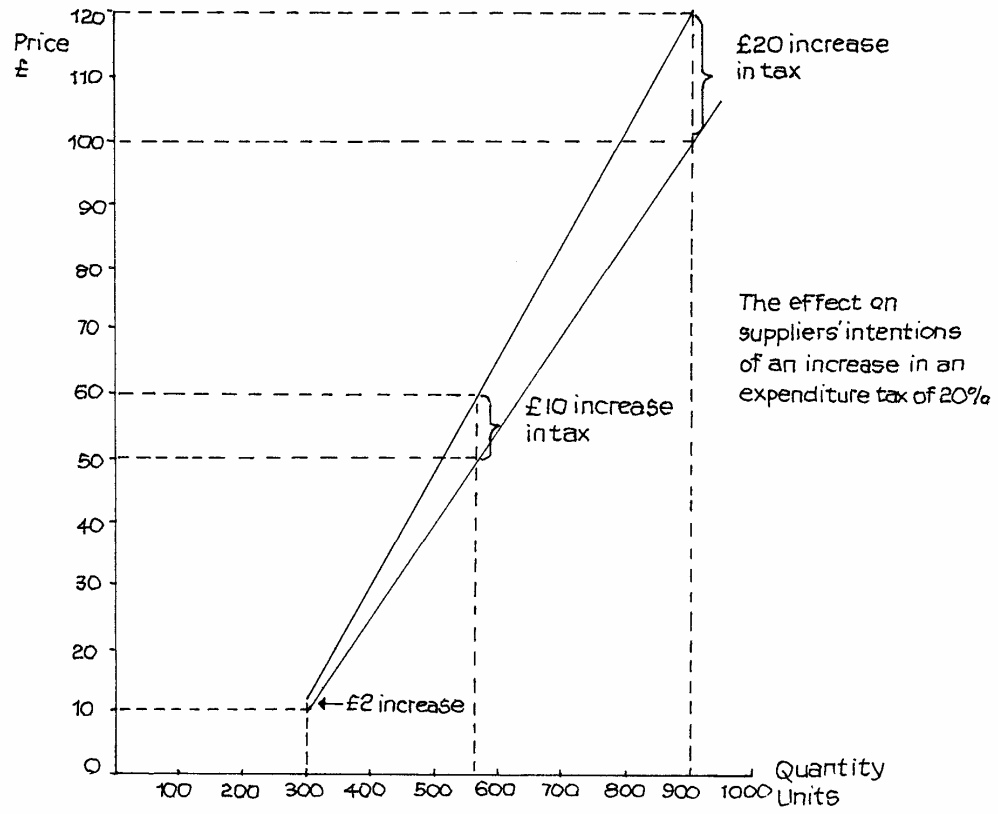


Figure 6.8: The effect on supply of an increase in an expenditure tax of 20%

Although suppliers will seek to recover the full amount of any additional expenditure tax from buyers there is no guarantee they will succeed in raising the price sufficiently to achieve this. The extent to which they can recover the tax or have to absorb it in their total costs through the more efficient use of their production resources depends largely on the strength of any price resistance shown by buyers. If buyers cease to buy the product at the increased price suppliers must reconsider their position. The possible consequences of this interaction between suppliers and buyers are examined later.

Effect of Tax on Price

We have just seen how the supply curve was likely to shift as a result of a change in an indirect tax or subsidy. For the likely effect on market price however, it is also necessary to take account of the conditions of demand, since it is unlikely that the producer's efforts to recoup the tax by adding this to the price will leave the quantity demanded in the market unchanged.

Look now at Figure 6.9. Here we show the movement of the supply curve from SS to S_1S_1 (resulting from the increase in tax) and the demand curve D_eD_e. The equilibrium price moves up (from Op to Op_1) but by an amount less than the increase in tax. The amount supplied to the market falls from Oq to Oq_1 and the output/quantity fall is greater than the price rise.

© ABE and RRC

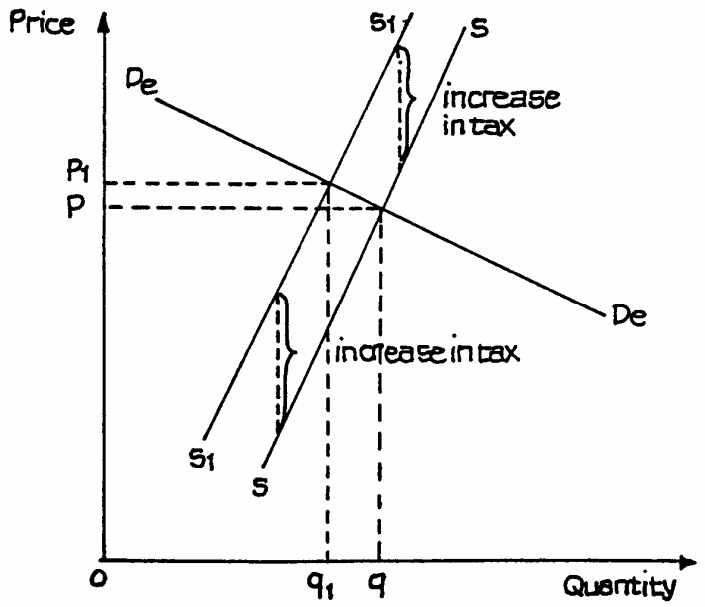


Figure 6.9: Effect of tax increase on supply and demand

Now look at Figure 6.10. Here we have the shift in supply curve SS to S_1S_1 and a demand curve D_1D_1. Again we have an increase in equilibrium price (Op to Op_1) and a reduction in quantity supplied (Oq to Oq_1). This time however, the reduction in quantity is less than the increase in price.

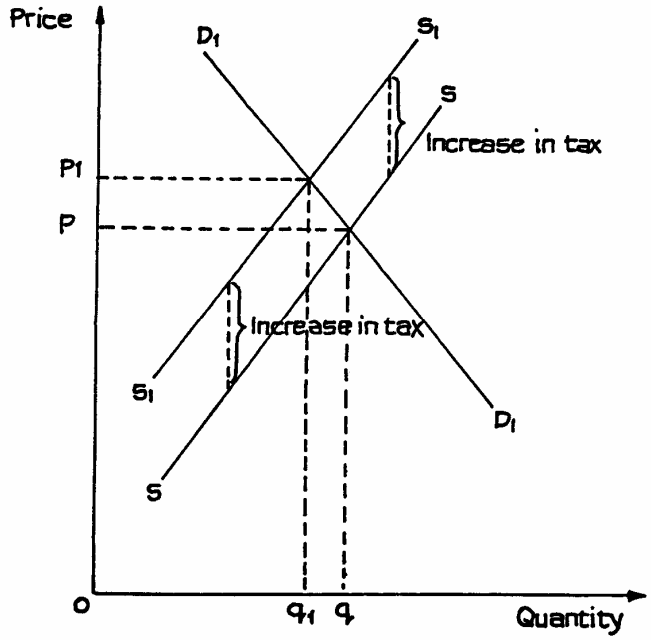


Figure 6.10: Effect of tax increase on supply and demand, price less elastic

Why the difference in the two situations? You will have noticed that the curve D_1D_1 is much steeper than D_eD_e. This reflects that demand in Figure 6.9 is more price elastic than demand in Figure 6.10. The two illustrations show that the more price elastic the demand for a product is, the smaller will be the market-price increase following an increase in indirect tax, and the greater will be the cutback in supply to the market.

This is after all really common sense. Price elasticity indicates the degree of responsiveness of quantity demanded to any change in price.

Subsidies

The effect of a subsidy will be the exact reverse of that of a tax. Instead of the movement of the supply curve from SS to S_1S_1, there is an increase in supply at all prices, i.e. as from S_1S_1 to SS, and there will be a reduction in market price, as from Op_1 to Op. Such a reduction is likely to have been the main government objective in arranging the subsidy, particularly if the good is a "socially worthy" one such as a basic food in a time of shortage, housing, or a merit good such as education or health care.

Remember also that the new supply curve need not be exactly parallel to the original before the tax or subsidy change. If the tax or subsidy increases with value, i.e. is an ad valorem tax or subsidy, the gap between the curves will increase as price rises, as illustrated in Figure 6.8.

Government Use of Indirect Taxes

If the government increases indirect tax on goods which are price elastic, it will not receive much extra tax but it will depress demand. If it imposes the tax on goods which are price inelastic, it will not have much effect on output but the government will collect more tax revenue.

If you now consider how price changes affect a person's pattern of expenditure and discretionary income you will realise that the effect of the tax may go further. Suppose there is a general increase in indirect tax on all goods. Some will be demand price inelastic, and their pricing will increase without much reduction in the amount supplied and bought. The buyers are paying more for nearly the same quantity of goods. This means they have less income to spend on other goods – they will have to cut purchases of goods which are price elastic.

The unfortunate producers of price-elastic goods will suffer a double blow. They will suffer a drop in demand from the tax increase and not be able to increase price by anything like the full amount of the tax, and they will suffer a further drop in demand because consumers' discretionary incomes have fallen. It is no surprise that business bankruptcies began to increase rapidly in the UK after a general increase in VAT.

We have so far assumed that these taxes would be used either to increase government revenues or to reduce consumer demand if the government believed that excess demand was causing inflation. There is however another aspect of government policy that is beginning to appear: this is the control of pollution, now recognised as a significant problem.

An indirect tax on expenditure could be used as an instrument to reduce demand, and hence the production or use of something that was believed to be a source of pollution. An example would be an additional tax on petrol to discourage the use of motor vehicles. However, as the demand for petrol is price inelastic then the tax will not have much effect on vehicle use but will reduce consumer incomes available for spending on other goods. One of the main reasons why demand for petrol for car use is price inelastic is because of the lack of satisfactory substitutes. As motor vehicle ownership has increased the demand for and supply of public transport has fallen; and as public transport provision falls and its price rises, so even more people are induced to use their own private cars.

We therefore conclude that a "pollution tax" on petrol would fail in its objective unless the government also made provision for (and probably subsidised) alternative public transport, at least in urban areas where cars are used for travel to work and for relatively short journeys. If the government also wished to discourage car use for longer journeys it would need to provide alternatives, probably in the form of subsidised rail travel combined with local transport to convey people from the main railheads. A tax is a very blunt instrument, and a government wishing to influence consumer behaviour needs to take many aspects into account. It is not sufficient simply to increase the price of the good whose use it wishes to discourage.

© ABE and RRC

Reverting to our general discussion of the effects of taxes on prices we have not taken into account differing elasticities of supply. This is because supply reactions will take place over a period of time. If suppliers can react by cutting back supply fairly quickly, then there will be further effects on market price. You can examine these for yourself by changing the supply curve to make it more elastic in Figures 6.9 and 6.10.

H. USING INDIRECT TAXES AND SUBSIDIES TO CORRECT MARKET DEFECTS

In this section we consolidate the preceding explanation and analysis by looking at how a government can use indirect taxes and subsidies to correct the market failures that result from externalities, the underconsumption of merit goods and the over-consumption of demerit goods.

If the consumption of a good or service is associated with a positive externality the demand curve for the good will fail to take this into account, and will only reflect the private benefits enjoyed by consumers. That is, individuals only consider their private benefit from consuming the good and the market demand curve measures the marginal private benefit derived from the good. In this case, because of the positive social benefit, the marginal social benefit curve will lie to the right of the demand curve. Such a good is a merit good and the position of the two curves is shown in Figure 6.11.

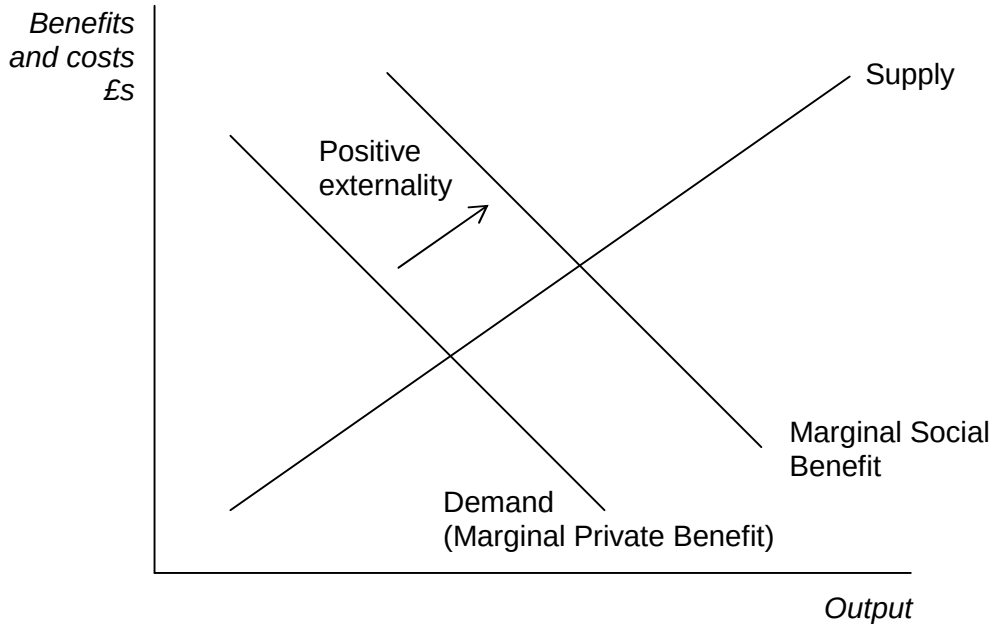


Figure 6.11:

Conversely, if the good is a demerit good, its marginal social benefit curve will lie to the left of its demand curve because of its negative externality. This is shown in Figure 6.12.**Error!**

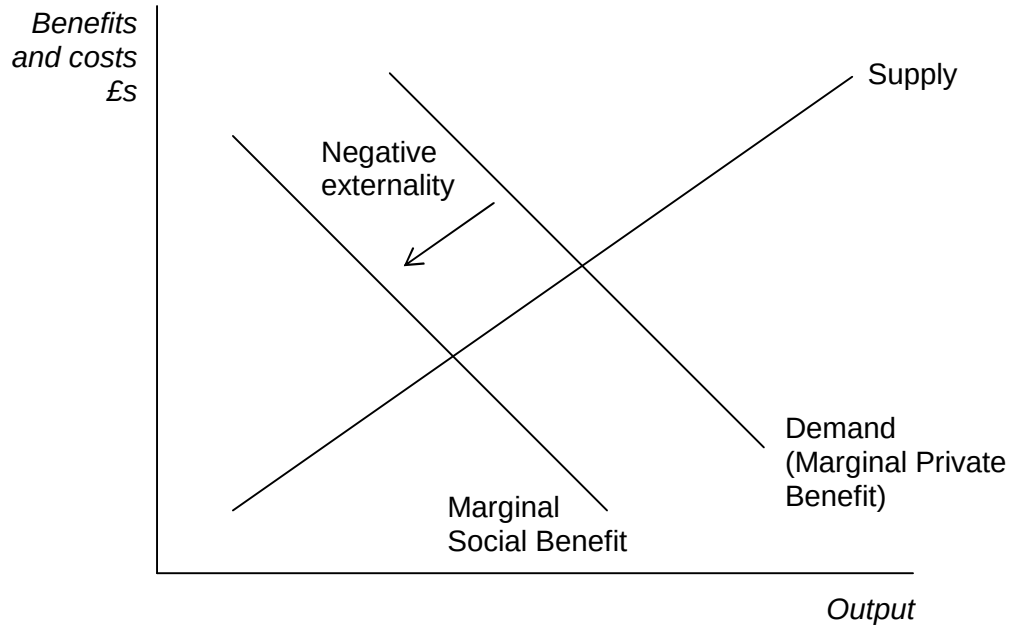


Figure 6.12:

A similar situation prevails with the negative externalities that can arise with production. The supply curve for a good or service only takes account of the private costs incurred by the producer of the good. The social costs created by any negative externalities during the process of production, such as water or atmospheric pollution, are ignored by the firm. In this case the firm's supply curve, which measures the marginal private cost of production, lies below the marginal social cost curve that adds the cost of the negative externality to the private costs. This is shown in Figure 6.13.

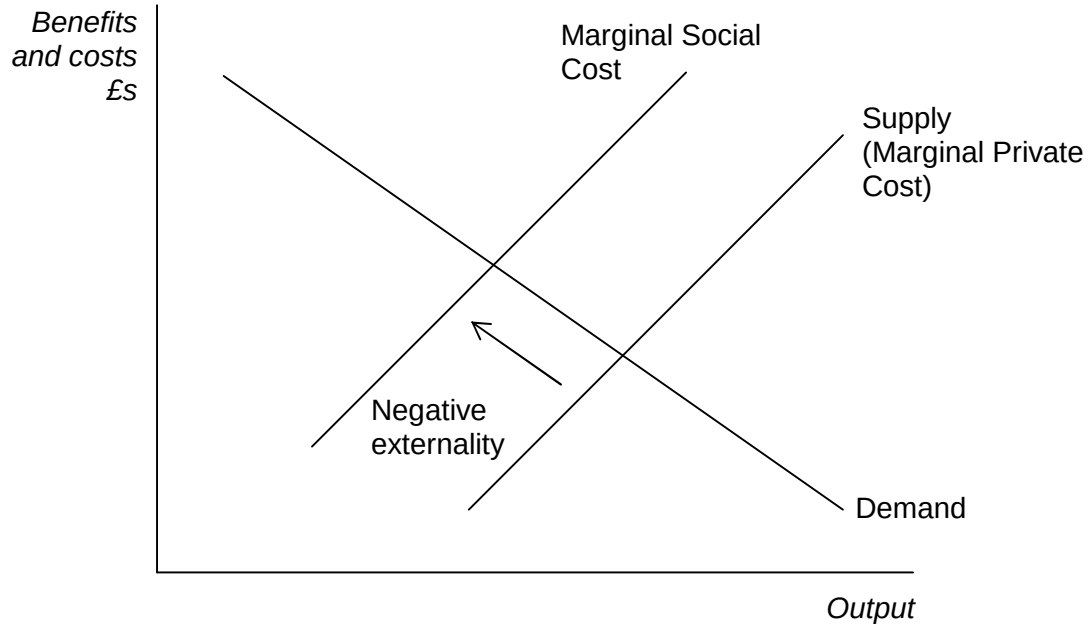


Figure 6.13:

In some cases the production process for a good or service creates a positive externality and the firm's supply curve fails to reflect the social cost of producing the good. For example, the smelting of aluminium involves large amounts of energy and creates waste heat. In the UAE the waste heat from the aluminium plants is used to distil sea water into fresh water that is

© ABE and RRC

then used for irrigation. Unfortunately such examples of positive externalities in production are much less common than the negative externalities due to pollution. Figure 6.14 illustrates the situation in which production creates a positive externality and the marginal social cost curve lies below the supply curve.

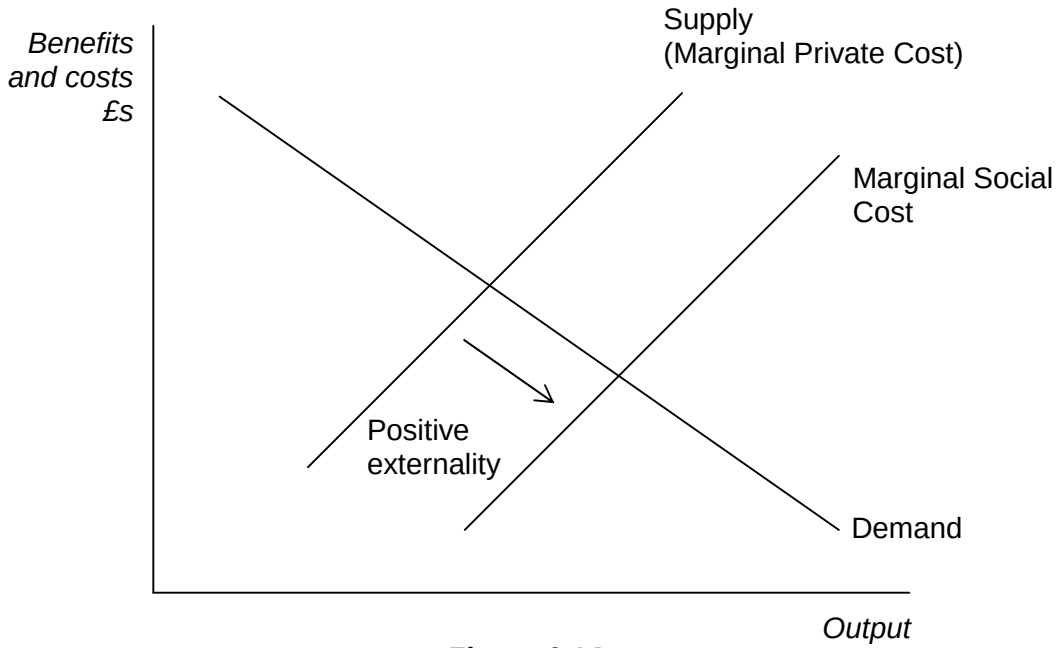


Figure 6.14:

Now we can combine the curves shown here and analyse the action required from government to correct the market failures that result from externalities in production and consumption. Figure 6.15 illustrates how a subsidy can be introduced when the marginal social benefit from a good exceeds the marginal private benefit. In the absence of a government subsidy, the free market equilibrium is where the demand and supply curves, which are also the marginal private benefit and cost curves, intersect at point E. At this point too little is being produced and consumed when account is taken of the marginal social benefits. The private market equilibrium quantity is Q_1 which is less than the socially optimum level of output Q_2, determined at the point where the marginal private and social costs are equal, point G.

© ABE and RRC

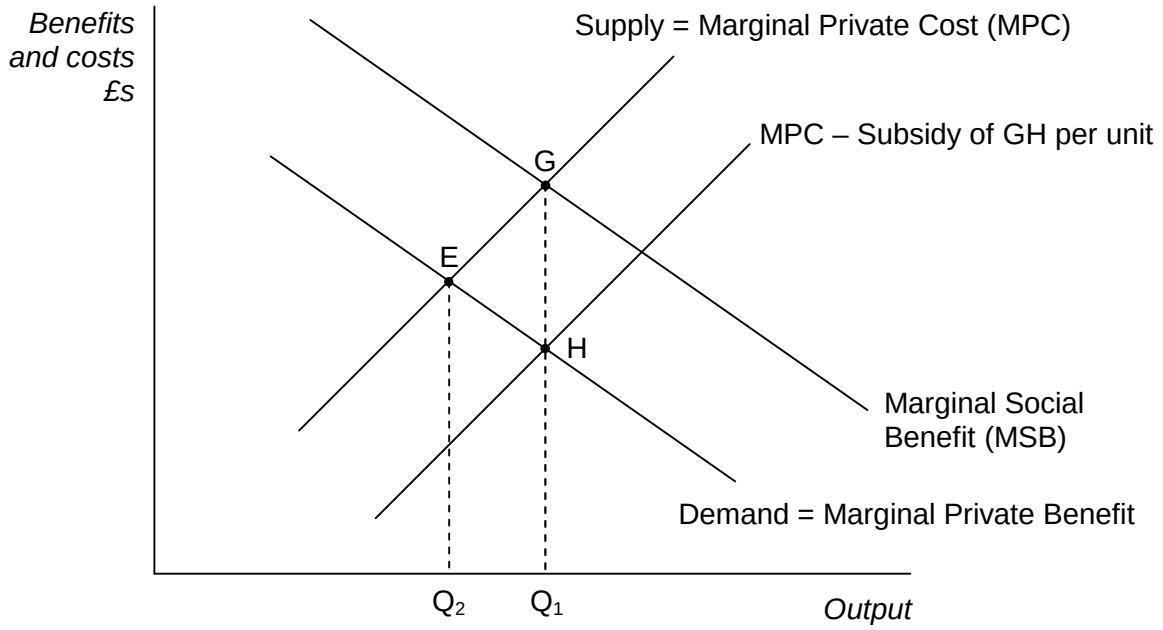


Figure 6.15:

To achieve the socially optimum level of production and consumption (Q_2), where the marginal social benefit equals the marginal private cost of production at G, the government should pay firms a production subsidy of GH per unit produced. The subsidy is equal to the value of the externality which is the difference between the marginal social and marginal private benefits at point G.

In the situation where there is a negative externality in production, because the marginal social cost of production exceeds the marginal private cost, firms overproduce the good in relation to the socially optimum level of production and consumption. Output, if left to the free market is Q_1, which exceeds the social optimum level of Q_2. To correct the market failure the government needs to make firms take account of the negative externality they are responsible for creating. The solution in this case is to impose an indirect tax on each unit of output equal to the difference between marginal social and private costs at the point where the marginal social cost curve intersects the marginal private benefit curve. This requires a tax of EF per unit. This is illustrated in Figure 16.16.

© ABE and RRC

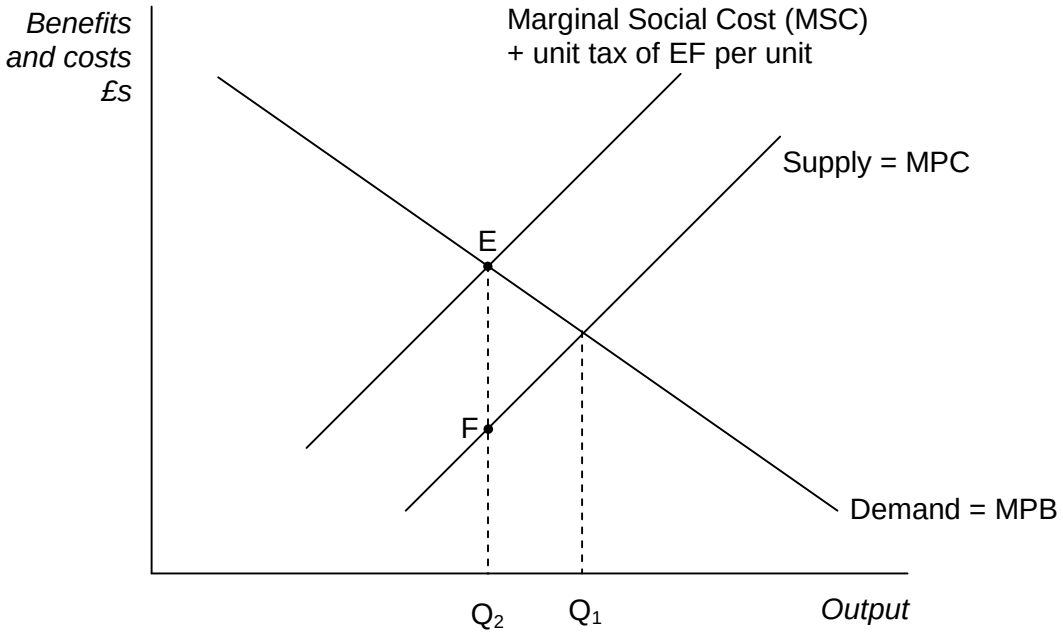


Figure 6.16:

Review Points

This is one of the most important units in the *Study Manual*. If you have not mastered its content you are unlikely to be able to achieve a satisfactory level of understanding of economics. Because of the fundamental role of the forces of supply and demand in the determination of prices in markets, and their significance for the behaviour of firms, and government intervention in markets, you need to make absolutely certain that you fully understand the content of this unit if you want to pass the examination in this subject.

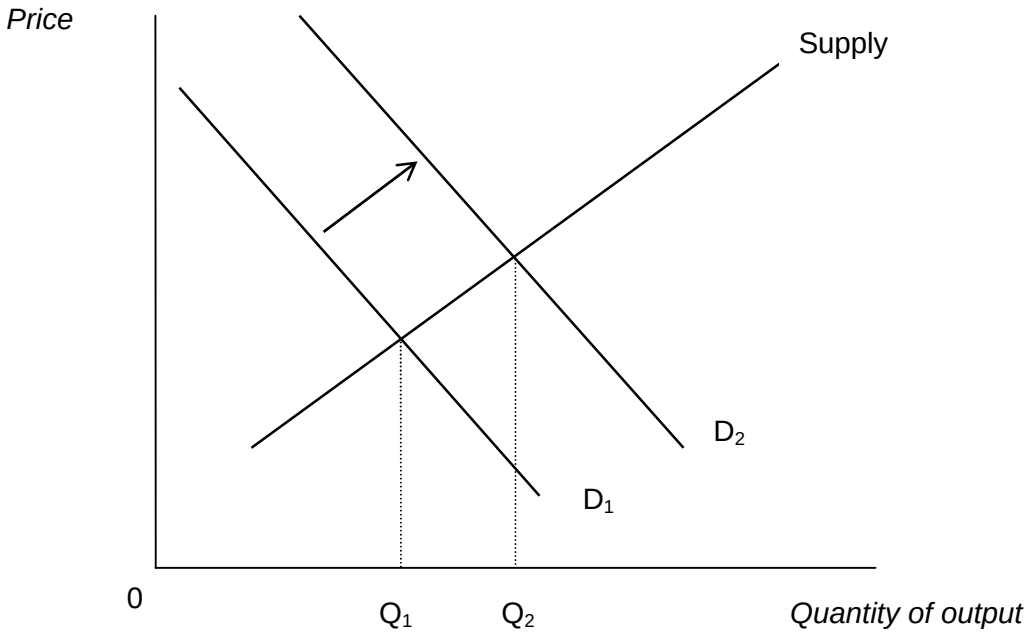
It is absolutely vital, before you continue with the next study unit, that you should go back to the start of this one and check that you have achieved the learning objectives and feel confident in undertaking demand and supply curve diagram analysis. If you do not think that you understand fully each of the learning outcomes you should spend more time reading the relevant sections.

You can test your understanding of what you have learnt, and your ability to use demand and supply curve analysis, by attempting to answer the following questions. Check all of your answers with the unit text.

1. In a free market, is the equilibrium market price determined by:
   (i) demand alone
   (ii) supply alone
   (iii) the interaction of demand and supply
   (iv) government intervention?
2. If the supply curve is upward sloping, other things remaining unchanged, will a rightward shift in market demand result in:
   (i) a decrease in the equilibrium price and quantity supplied, or
   (ii) an increase in the equilibrium price and quantity supplied?

© ABE and RRC

3. If the demand curve is downward sloping, other things remaining unchanged, will a rightward shift in the supply curve result in:
   (i) a decrease in the equilibrium price and an increase in the quantity supplied, or
   (ii) an increase in the equilibrium price and quantity supplied?
4. The following diagram shows the initial equilibrium position, Q_1, in the market for a normal good and a second demand curve D_2.

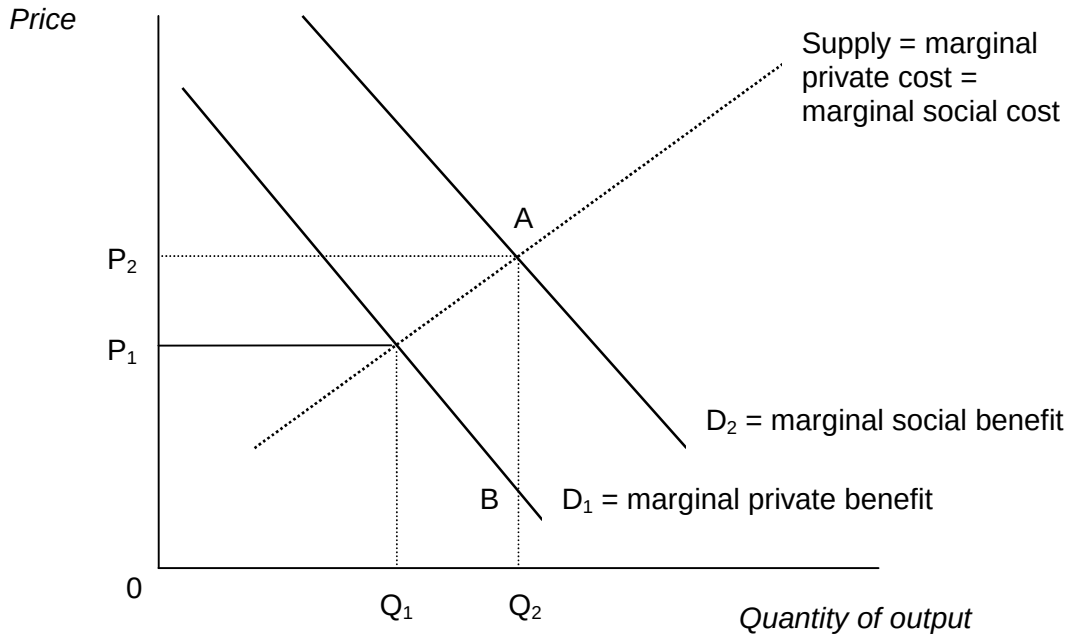


   Could the rightward shift in the demand curve be the result of:
   (i) a decrease in the price of a substitute good
   (ii) an increase in the incomes of consumers
   (iii) the introduction of an indirect tax on the good by the government?
5. Explain the meaning of the following:
   (i) externality
   (ii) social cost
   (iii) social benefit.
6. If consumption of a good yields a positive external benefit, is the good referred to as:
   (i) a demerit good, or
   (ii) a merit good?
7. In the absence of intervention by the government, if the social marginal cost of a good exceeds its marginal private cost is the good:
   (i) overproduced
   (ii) under-consumed?

© ABE and RRC

8. Explain the meaning of the following:
   (i) price floor
   (ii) price ceiling
   (iii) output quota.

9. The following diagram shows the free market equilibrium position, Q_1, for a merit good.



To achieve a socially optimal level of production and consumption of the good should the government intervene in the market and:

(i) pay producers a subsidy of AB per unit
(ii) tax producers AB per unit produced
(iii) impose a price ceiling of P_2
(iv) impose a price floor of P_1?

© ABE and RRC

Study Unit 7

Market Structures: Perfect Competition versus Monopoly

Contents ***Page***

A. Meaning and Importance of Competition 130

B. Perfect Competition 131
Definition 131
Conditions for Perfect Competition 132
Movement towards Equilibrium in Perfectly Competitive Markets 133
Views on Perfect Competition 136
Profit Maximisation as a Result of Perfect Competition 136

C. Monopoly 137
Definition 137
Sources of Monopoly 137
The Monopoly Model 138
Is Monopoly Good? 139

© ABE and RRC

Objectives

The aim of this unit is to: explain the profit-maximising outcomes under monopoly and perfect competition in the short and long run; identify the differences between the two market structures; examine the effects of changes in government policy upon these markets.

When you have completed this study unit you will be able to:

- identify, using diagrams, the characteristics of perfect competition at the firm and industry level and identify, in numerical and/or diagrammatic examples, equilibrium price, firm and/or industry quantity, profit, marginal cost, average cost, marginal revenue and average revenue
- examine, for perfect competition, the effects of changes in the conditions of the industry upon the market equilibrium in the short and long run and discuss the mechanism by which the industry moves from the short-run to the long-run equilibrium and discuss the welfare implications of perfect competition
- identify, using diagrams, the characteristics of monopoly and explain the relationship between average and marginal revenue, and identify, in numerical and/or diagrammatic examples, equilibrium price, output, profit, total cost, total revenue, marginal cost, average cost, marginal revenue and deadweight loss
- examine, for monopoly, the effects of changes in the conditions of the industry upon the market equilibrium in the short and long run and discuss the welfare implications of monopoly with reference to the deadweight loss triangle and X-inefficiency
- discuss the merits of policy alternatives aimed at reducing the social cost of monopoly
- solve basic diagrammatic and numerical problems under monopoly and perfect competition
- identify and discuss real world examples of industries with similar characteristics to the models of perfect competition and monopoly.

A. MEANING AND IMPORTANCE OF COMPETITION

"Competition" is one of those simple words which are common in everyday speech. We all assume we understand what it means, but when we try and explain it, it starts to present problems. Ask yourself what benefits you think you get from competition as a consumer. Suppose you think in terms of being able to buy from different suppliers, and being able to choose from a variety of different but broadly similar goods – for example choosing shoes of different styles, sizes, quality and price ranges. Perhaps you think of having some power as a consumer to bargain over price, or awareness that some suppliers will charge lower prices than others. Notice that the word that recurs constantly when most of us think about competition is *choice*. You and I, as consumers, value the ability to choose between a range of goods, different prices and different standards of quality and service.

Because of the buyers' ability to choose and apply pressure on prices, we expect competition to oblige producers and distributors to use their resources efficiently and keep production and distribution costs low. Competition is usually thought to be a very powerful force to ensure production efficiency.

Competition is thus widely believed to be a desirable feature of markets. Most of the major modern market economies have legislation and institutions concerned with preserving or increasing competition. The Treaty of Rome, the founding treaty of the European Economic Community – now the European Union – contains a strong commitment to competition and the prevention of attempts to limit it.

© ABE and RRC

Economists have generally been in favour of competition as a force likely to increase the efficient use of scarce resources, and they have developed a concept of *perfect competition* which we shall examine in this study unit. However more recently they have recognised that traditional views of competition have limitations, and that the pressures on business firms are more complex than have sometimes been believed in the past. There is also a recognition that increased competition can sometimes have consequences that are not beneficial to consumers, or which are not socially very desirable. In particular, competition may be harmful, or at least lead to a socially suboptimal outcome, if firms take no account of the existence of positive and negative externalities in production, and their impact on the environment.

So we must be careful in our assessment of the benefits of competition, and be prepared to be critical when examining some of the traditional economic models of competitive markets. These models have been developed in the belief that the degree of competition in a market is likely to influence the behaviour and performance of firms operating in it. In this study unit we look at some of the best known models; these provide an essential starting point for understanding the often complex markets existing in modern economies. However, we must be equally careful in our assessment of competition that we do not impose our values of what is good or bad for society on others who may have different values. It is also important to note the influence of technology on markets and competition. For example, the rapid growth of modern low cost communications and knowledge sharing in the form of mobile phones and the Internet have significantly increased competition, both in markets within countries and between countries. Indeed, the Internet has made the economists' ideal model of perfect competition a much more real description of how many markets now work in the real world.

B. PERFECT COMPETITION

Definition

Our first theoretical model covers the situation where the economic market operates in its purest or most perfect form. *Perfect competition is the state of affairs existing in a market totally free from imperfections in the communication and interaction of the economic forces of supply and demand.*

Some writers like to make a distinction between *perfect or ideal markets* and *perfect competition*, in addition to the distinction between the market as an area and competition as a condition found in that area. They suggest that the conditions for perfect competition are satisfied when the individual firm is a "price-taker", i.e. when it can sell all that it can produce at the market price, which by itself it cannot alter, and when buyers are indifferent as to which seller's product they buy at that price. Such a very limited set of requirements would be satisfied when firms in an industry were subject to a regulated price set by a government or some other regulatory body which had powers to buy goods unsaleable in the market. This would certainly not be a perfect market.

For true perfect competition to exist, it seems more realistic to stipulate that sellers must be free to enter and leave the market, so that total supply can change and bring about the equilibrium position. Just to establish a market price through some form of price regulation would not produce the same result, unless the regulating body is very sensitive to demand shifts, and production plans can be adapted quickly.

So it seems that full operation of *perfect competition* can be achieved only in a *perfect economic market*, and to put too much emphasis on differences between the two does not really help very much in our analysis of the main market forces.

© ABE and RRC

Conditions for Perfect Competition

These can be summarised as follows:

(a) Goods must be Homogeneous

This means that in the *perception* of the buyer, all units of the goods offered by all suppliers are equally acceptable. The buyer is indifferent as to which unit he or she receives, as long as it conforms to any description adopted by, and understood in, the market.

Notice that it is the perception of the buyer that is important. Suppose two large retail stores make an arrangement with a manufacturer to be supplied with canned baked beans in plain tins. The manufacturer supplies beans of the same type and quality to each retailer in the plain cans quite impartially. However, each store adds its own label to the cans and sells the beans under completely different brand names and at slightly different prices. The products are physically the same, but they are not homogeneous, because the public perceives them as different and competing products.

(b) Perfect Transport and Communications

All consumers in the market must have the same information. Suppliers must have access to the same information about production factors and the technical conditions of production. No producer is in a more favoured situation than any other.

(c) Price Established Only by Market Forces

No producer and no buyer is able to influence the price by his or her own actions, nor by actions agreed with other producers or buyers. There is no degree of monopoly power in the market.

(d) Economic Motives Only

The actions of suppliers and buyers are influenced only by economic motives. If buyers or sellers are influenced by a desire to support a charity or a political party the market will not be purely economic, however worthy the social motives.

Economic rationality in a market economy assumes an underlying self-interest and a desire to maximise benefits that can be gained from available scarce resources. For the consumer this means maximising utility, as defined in Study Unit 2, while for producers it is usually interpreted as wishing to maximise profit – an objective examined later.

(e) No Barriers Limiting Market Entry and Exit

Suppliers and buyers must be free to enter and leave the market as they choose and as they are guided by considerations of profit and utility. This is a very important element in any competitive market and in some modern models of market behaviour, notably that of *contestable markets*, it is the most important consideration.

Barriers to market entry and exit may be "natural", i.e. arising out of the nature of the goods or the production process, or "artificial", i.e. arising out of market regulations.

Natural barriers are highest when production requires large amounts of highly specialised capital, e.g. oil exploration and extraction or motor vehicle assembly. Only firms with access to very large amounts of finance can enter these markets. Once this capital has been acquired, the firms are committed to staying in the market, since exit would usually involve very large financial losses. Natural barriers are low when little specialised capital or skill are needed to commence production.

When natural barriers are low established producers may seek to protect themselves from new entry by building *artificial barriers*. These barriers may be membership of a trade or professional association (entry to which may require a long period of

© ABE and RRC

apprenticeship), education or high membership fees. It is not unknown for established traders to prevent new entry illegally by the use of force, as in the case of ice cream selling in some areas and, of course, street trading in illegal drugs.

The lower the barriers, both natural and artificial, the more contestable the market; the theory of contestable markets suggests that contestability is a powerful force determining the behaviour of suppliers in a market. If producers know that they can easily be challenged by new competitors, they will behave as if they were subject to competition because they will not wish to provide incentives for new firms to come into the market. Such incentives would include supernormal profit or the existence of buyers who were dissatisfied with existing goods, standards of service or prices.

Consequently we would expect a perfectly contestable market to exhibit most if not all the characteristics of perfect competition.

Movement towards Equilibrium in Perfectly Competitive Markets

We can now examine the behaviour of firms operating under conditions of perfect competition.

If we assume that the firm is experiencing diminishing marginal returns and can sell all it can produce at the market price, over which it has no control, then it will have average and marginal cost curves and an average revenue curve as shown in Figure 7.1. Since all units of the good are sold at the same price whatever the firm's sales level, price will equal average revenue and will also be the same as marginal revenue.

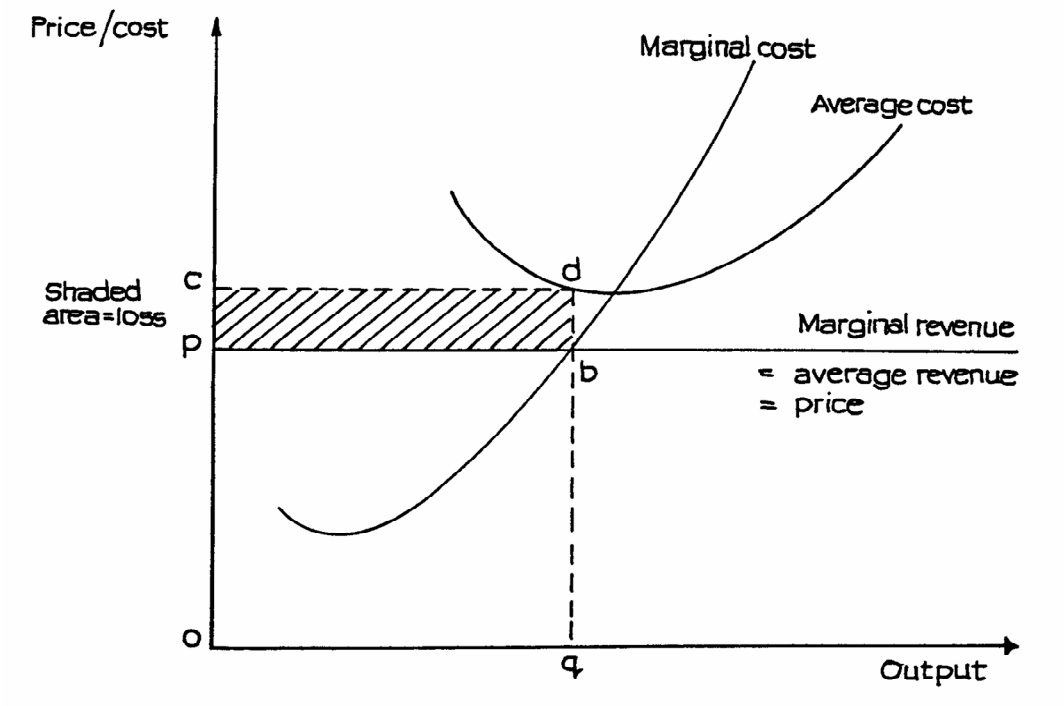


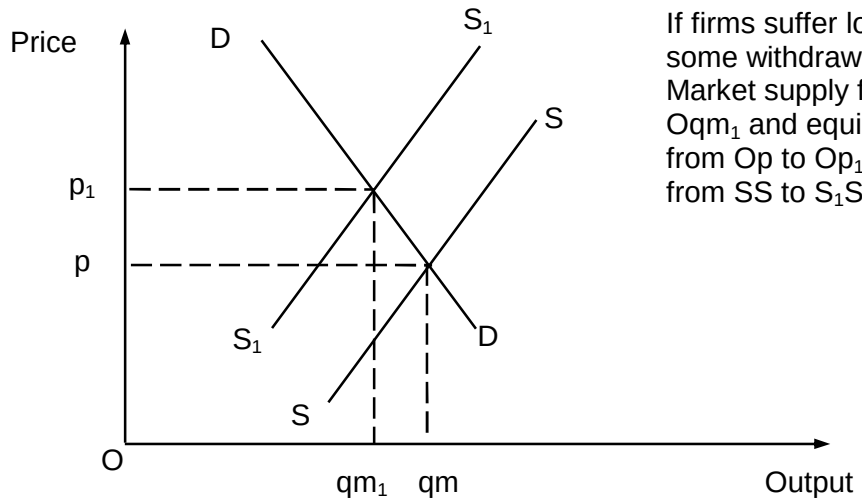
Figure 7.1: Marginal cost, average cost and marginal revenue

Suppose the price resulting from the interaction of supply and demand in the market as a whole is Op; then there is no level of output at which the firm can produce at a profit.

At all levels of output price, average revenue is below the average cost curve. However, the profit-maximising condition of marginal cost equals marginal revenue is also the loss-minimising condition, so the best output for the firm to choose is at Oq where marginal cost equals marginal revenue. At this output level, average cost at Oc is higher than average revenue at Op, so the firm suffers a loss equal to the shaded area (cdbp).

© ABE and RRC

Given the conditions for perfect competition, if this is the situation faced by one firm, it is the situation of all firms subject to the same market information and technology. Firms cannot continue indefinitely suffering losses. Some will withdraw from the market (remember that unrestricted entry and exit is another condition of this market) because they are less able to withstand losses or they have other markets they can enter. As supply declines, so the total market supply curve will move to the left, as shown in Figure 7.2.



If firms suffer losses at price Op_1 some withdraw from the market. Market supply falls from Oqm to Oqm_1 and equilibrium price rises from Op to Op_1 as supply shifts from SS to S_1S_1.

Figure 7.2: Market supply curve moves left

The market equilibrium price then rises – assuming that demand remains unchanged. Supposing the equilibrium price moves up from Op to Op_1, this produces the situation for the individual firm illustrated by Figure 7.3.

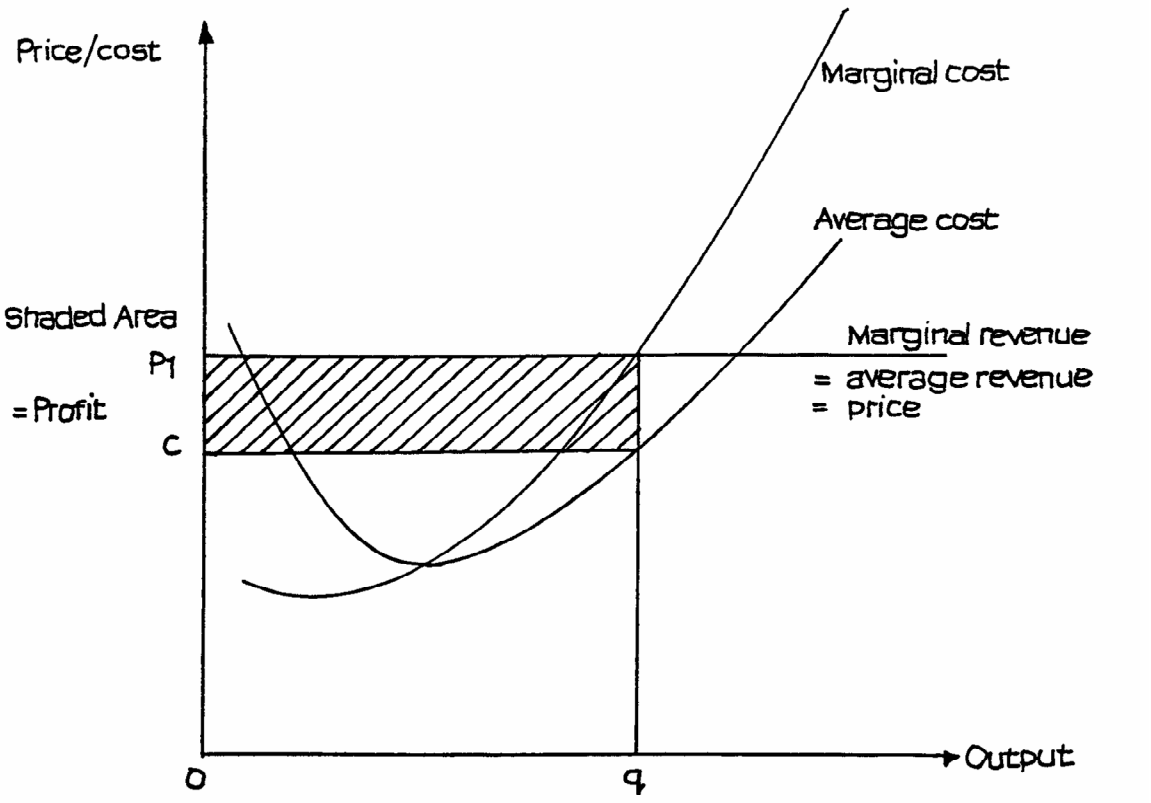


Figure 7.3: Market equilibrium price rises

Now we see that the average revenue at Op_1 is higher than average cost at Oc, and the firm is enjoying profits, represented by the shaded area. Notice that once again the most profitable output to aim at is at Oq, where marginal cost is just equal to marginal revenue.

Now, given our earlier assumptions, all firms are making profits. If we have defined cost to include a normal return to *all production factors* (including some return to enterprise in the form of a minimum profit to keep firms in the market and provide necessary capital investment) then this shaded area profit is an additional or abnormal profit, resulting only from the special market opportunities.

Owing to perfect communication and free entry, new firms will enter the market to take advantage of these profits. Supply will now increase – the supply curve will move to the right and equilibrium price will fall.

Suppose it falls to a position between Op and Op_1, say to Op_e where price/average revenue is just equal to average cost. Now the individual firm is in the position illustrated in Figure 7.4. Here, there is neither abnormal profit nor loss. We assume that the firm's costs include an element of normal profit, which can be defined as a fair return to the firm's enterprise, or sometimes as that amount of profit which is sufficient to keep firms operating in that market. This normal profit is included therefore in the average cost curve. There is no incentive for firms to move into or out of the market: there is no reason why supply should shift – and, as long as demand remains unchanged, there is no reason for any movement in this equilibrium balance.

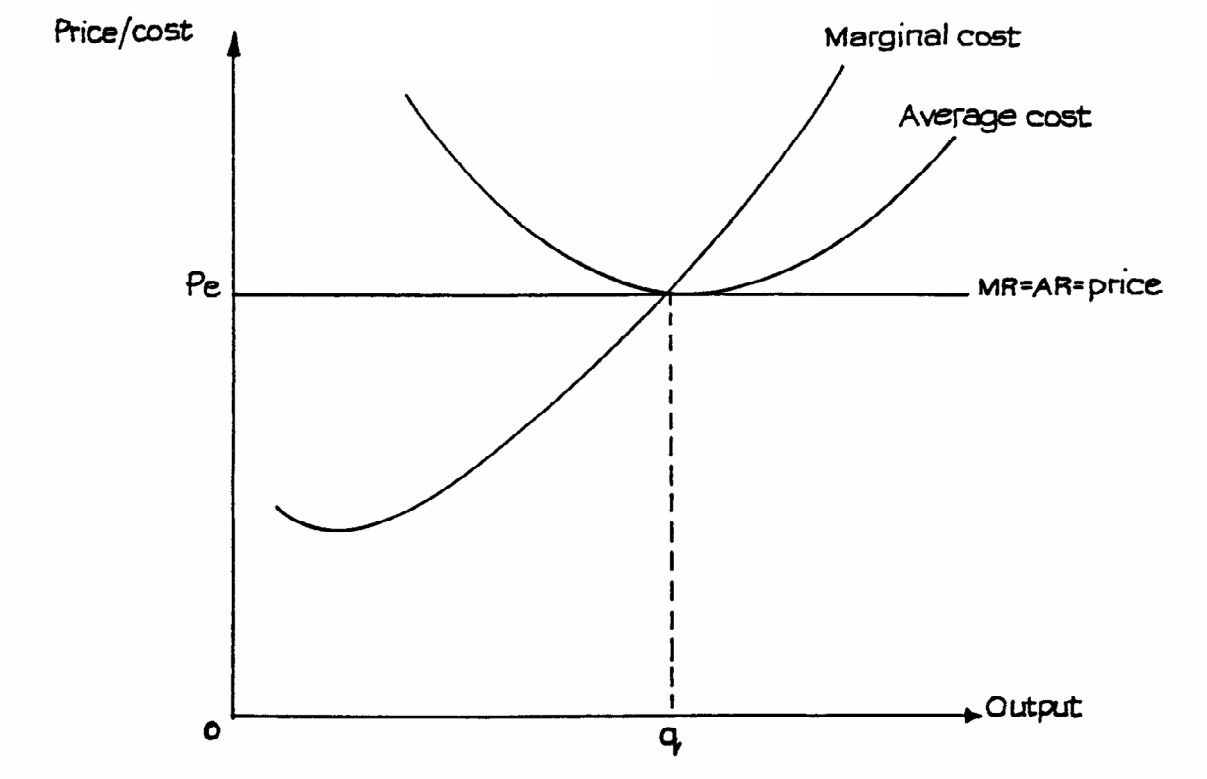


Figure 7.4: Perfect competition

It is on the basis of this kind of argument that textbooks and examiners sometimes make much of the distinction between *short-run equilibrium* in perfect competition where abnormal profits or losses can be experienced, and *long-run equilibrium* where only "normal" profits (included in the average total cost curve) are possible. However, we should stress that these are really only partial equilibrium positions relating to supply alone. The model says nothing about influences on demand which is often far from stable. A shift in demand will be quickly reflected in a shift in supply to readjust output to the new market price. Consequently, in

© ABE and RRC

markets where demand is inherently unstable – as in the stock and commodity exchanges – long-run equilibrium may never be reached as suppliers are constantly adapting to the shifting market environment.

Views on Perfect Competition

Economists often favour perfect competition on the following grounds:

- The elimination of abnormal profit, as shown in Figure 7.4 (compare to Figure 7.3).
- Efficient use of resources. Notice that, in equilibrium, the bringing together of price and marginal cost and the elimination of abnormal profit means that producers will produce when the average cost curve is at its lowest point (where marginal cost equals average cost). There is then a tendency to encourage producers to reduce average costs as much as possible. This is equivalent to making the most efficient use of resources.
- Price is equal to marginal cost. Price is the money value of the utility gained by the last or marginal consumer, i.e. *marginal utility*. When marginal cost equals marginal utility, as in perfect competition, the cost of producing the last unit is just equal to the value of the utility given by that unit to its consumer. If this were true in all cases, then the total cost of production would equal the total value of utility received. It is suggested this would be the best possible use of all resources.

Not everyone accepts these arguments, and you should consider the contents of this section in conjunction with the discussion of monopoly, later.

One of the arguments against perfect competition is that it prevents producers from making the profit necessary to provide funds for investment and research, to find better ways of producing goods. Another argument is that competition can be wasteful, as resources are doing the same things. If there were fewer competing firms, total costs could be reduced and some resources freed to produce something else.

Firms dislike perfect competition because, as indicated earlier, prices are unstable. If communications are good, then supply can adapt very quickly to price changes caused by changes in demand. The result is that prices are constantly adapting to new equilibrium positions – as with the Stock Exchange, which is still the common textbook example of a market which is close to perfect competition. In the Stock Exchange, prices change daily, and even hourly.

Manufacturers cannot tolerate swiftly-moving prices like this – they could survive in such a market only if they could keep changing the prices paid for production factors, including the wages paid to workers. Trade unions have sought to achieve stable jobs – and, preferably, rising wages. Producers then want stable – and, preferably, rising – prices. Those economists who argue for perfect competition in the consumer interest, and then argue for stable wages and secure employment, are being illogical. These two conditions cannot exist together.

So perfect competition may or may not be ideal from a purely economic viewpoint. It is certainly far from ideal from a social standpoint.

Profit Maximisation as a Result of Perfect Competition

Notice that the only output enabling the firm to survive in the equilibrium condition illustrated in Figure 7.4 is where marginal cost equals marginal revenue. The removal of abnormal profit ensures that the average cost curve is at a tangent to the average revenue curve, and as this is horizontal, it follows that the average cost curve must be at a tangent at its lowest point, i.e. where average cost equals marginal cost.

This is what is meant by saying profit maximisation is a survival condition resulting from perfect competition. Only by achieving this profit-maximising output can the individual firm

© ABE and RRC

avoid losses. Whether it achieves this intentionally or by trial and error does not matter; failure to achieve it means eventual failure to exist in the market.

C. MONOPOLY

Definition

Monopoly is the opposite extreme to perfect competition. It exists when there is only one supplier for a particular product and there are no close substitutes for that product.

Again, we have to be careful how we define the product. For example, the Post Office has a monopoly in the delivery of low-price letter mail in Britain. However it does not have a monopoly in personal and business communication, and in recent years the volume of letter mail has declined in the face of competition from the telephone, fax and from private firms of leaflet distributors. It now faces more competition from email and Internet services. Historically almost all monopolies are subject to destruction by the onward march of technology.

Sources of Monopoly

Monopoly can arise in three ways: by operation of the law, by possession of a unique feature, or by the achievement of market control.

(a) Operation of Law

This is a very old source of monopoly power. Kings used to sell monopolies in Europe to raise money: they sold people the right to be sole suppliers of a necessary product, such as salt, in a given area. The monopolist could rely on the support of the King's officers to protect his monopoly, and the profits he could make more than covered the fee he had to pay for his position.

Today, some countries may grant a company the right to be sole supplier of a product or service (e.g. telephones) in return for some measure of State inspection and control over profits and prices. In Britain, before 1979, it was usual for such monopolies to be public corporations under public ownership and control. This has been changed by the privatisation programme, which has resulted in a policy of separating regulation from operation. Some important public utilities are now legally companies in the private sector (e.g. British Telecom and British Gas), but are subject to government influence as a shareholder, and regulation by separate bodies (OFTEL and OFGEM respectively). (OFGEM is also the electricity regulator and water industries are regulated by OFWAT.)

A more limited monopoly power is granted under patent and copyright laws, which are similar in most countries. The idea of a patent is that the inventor of a new idea shares his or her knowledge with the State for the public benefit, in return for a monopoly control over the use of the idea for a limited number of years. If rival suppliers are unable to develop a competing product without breaking the patent, this form of monopoly can be very valuable – for example the monopoly enjoyed for some years by the Polaroid instant film-developing process.

(b) Possession of a Unique Feature

Individuals have monopoly control over the supply of their own skills, and this may be a source of considerable profit. The top footballers, tennis players and entertainers are monopolists of this type. When the skill lies in producing something written or recorded, then the monopoly position is protected by copyright laws – which, however, modern technology has made more difficult to enforce.

© ABE and RRC

(c) Market Control

It is difficult to achieve total monopoly over supply without the protection of the law, although it is not unknown – especially in the production of some intermediate products. For a number of years, all the valves for pneumatic tyres on British motor vehicles were produced by one manufacturer. Such a monopoly rarely lasts very long. When a large rival decides to challenge the monopolist, there is little that can be done to prevent this.

The Monopoly Model

The model has been developed to explain the outcome of a monopoly not subject to any special legal protection or control. It assumes that the firm is pursuing a profit-maximising objective, and that it is able to make abnormal profits.

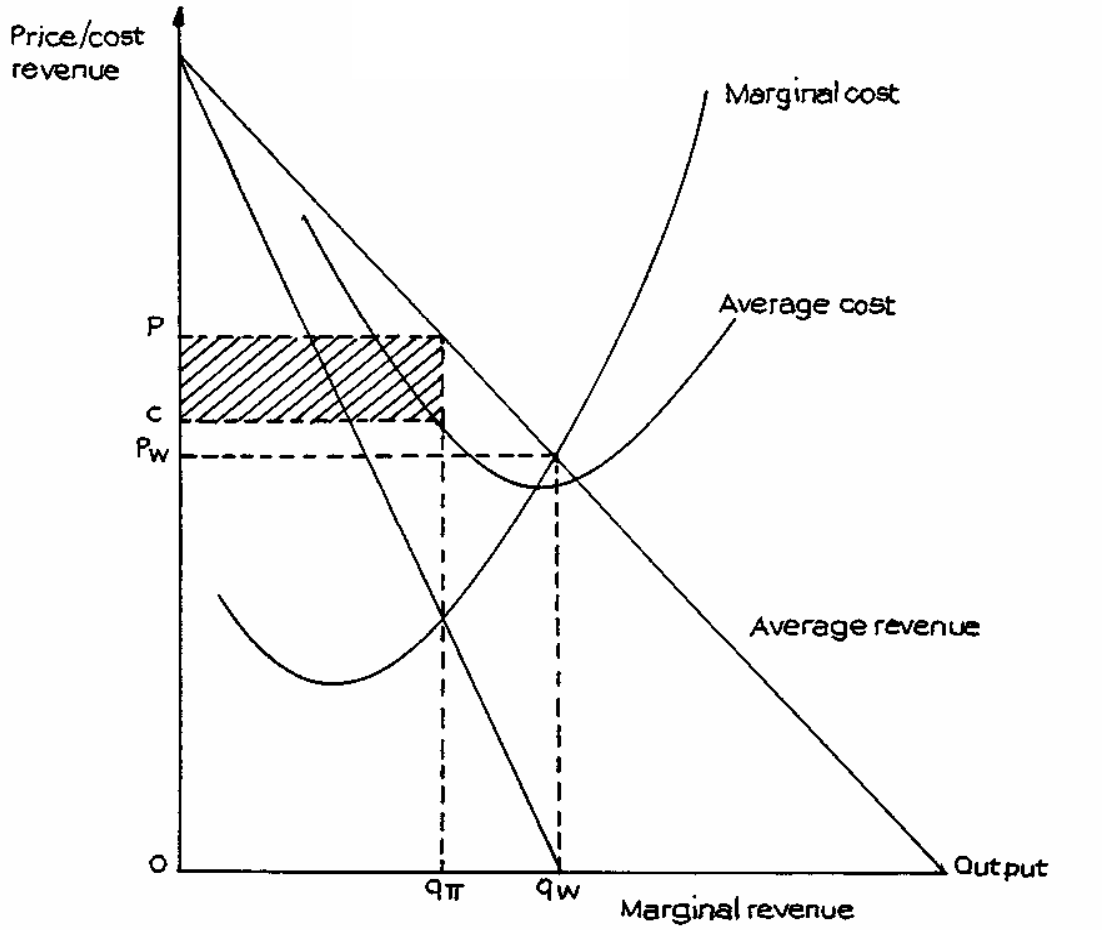


Figure 7.5 : Monopoly

A monopolist's output is the total market supply, and the demand for its product is the total market demand. The firm will thus face a downward-sloping demand curve. If we assume that it is not practising price discrimination, then this curve will be the price/average revenue curve. The graphical model is shown in Figure 7.5.

The profit-maximising monopolist will produce at output Oq_{π}, where marginal cost equals marginal revenue, and will charge price Op. Abnormal profit is represented by the shaded area. The average cost is Oc, so Op – Oc is the average profit earned on each unit of product sold.

© ABE and RRC

If the firm were to set price to equal marginal cost, which is the position desirable from the consumer viewpoint, it would produce output Oq_w and charge the lower price Op_w. This is why the profit-maximising monopolist is said to restrict output and increase price in comparison with a firm operating in a competitive market. Is it the case that monopoly is worse than competition and operates against the public interest?

Is Monopoly Good?

There is much evidence that large firms with considerable market power may *not* maximise profits but may pursue quite different objectives, such as growth or sales revenue maximisation. The average cost curve was drawn on the basis that abnormal profit was being made. There is nothing in the model itself that says that the average cost curve must be this shape and in this position. We can move it up or down without affecting the other curves, and so alter the profit quite legitimately.

In short, the model *proves* nothing. It simply illustrates the assumptions made. Notice that, if we drop the profit-maximising requirement, we can allow the firm to increase output and reduce price, and so come closer to the consumer-benefiting output level of Oqw. This would also reduce average cost and allow the firm to make more efficient use of its resources.

In answer to the charge that monopoly is against the public interest because it restricts output and raises price, the following arguments are often put forward in defence of monopoly:

(a) The monopolist's size and ability to produce for the whole market enables it to achieve economies of scale, so that costs are actually lower than they would be under perfect competition.

(b) The monopolist employs professional managers who make more efficient use of available resources than small owner/managers, who often lack managerial skill.

(c) The monopolist does not maximise profits but is content with just a satisfactory level of profit.

(d) Some element of abnormal or monopoly profit (normal profit is considered to be included in the firm's costs as for perfect competition) is desirable, so that the firm can:

  (i) spend money on research and gather funds for further capital investment;

  (ii) have the incentive to take risks and innovate, and sometimes suffer losses that would cripple smaller firms.

The position where a monopolist is actually able to charge lower prices than would be possible under perfect competition is illustrated in Figure 7.6. Here, for simplicity, constant average total costs have been assumed and the monopolist's cost curve is below that of small firms by reason of economies of scale and improved technology. Assuming that the monopolist seeks to maximise profits, the appropriate price will be Pm, still higher than the perfectly competitive price of Pc. However, this could be reduced if the monopolist had some other objective such as maximising growth or revenue. The revenue-maximising price (Pr), i.e. the price applicable to producing at the quantity level where marginal revenue is zero, and therefore total revenue is at its maximum, is lower than the perfectly competitive price of Pc. Notice that, unlike the firm under perfect competition, the monopolist can charge a range of prices, depending upon the firm's objectives, and still make a profit.

© ABE and RRC

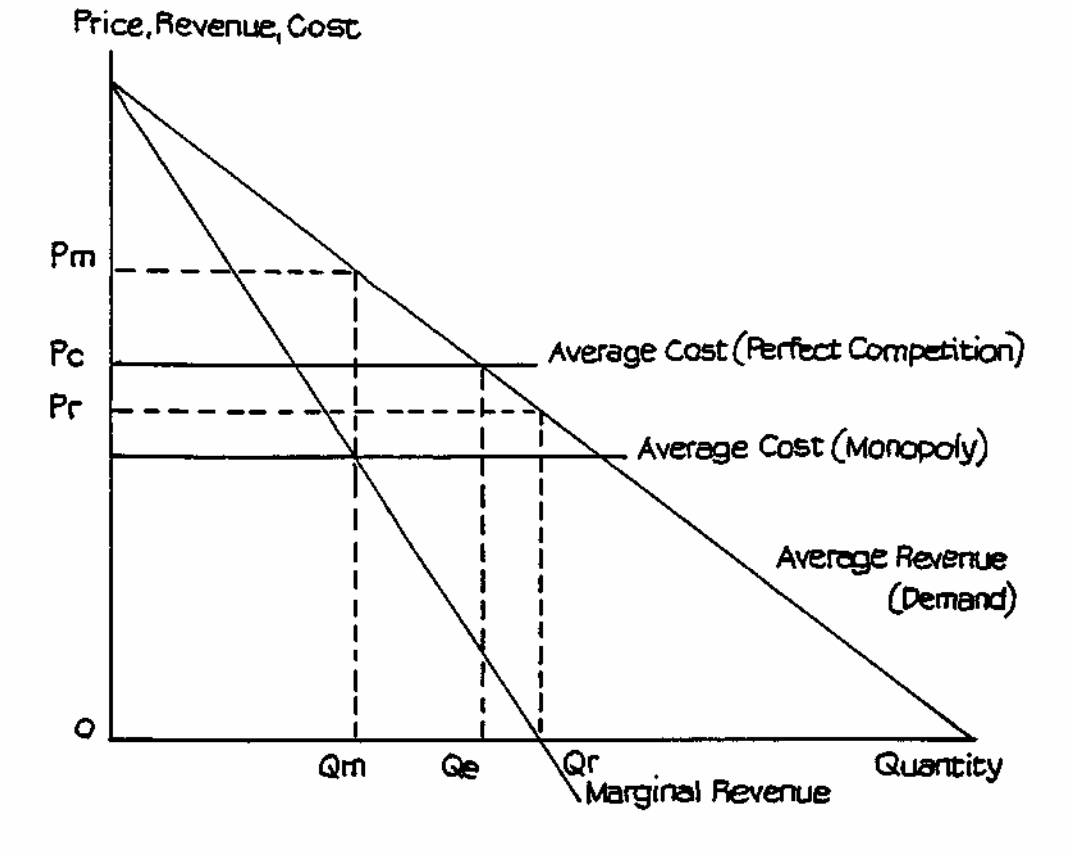


Figure 7.6: Price and output under perfect competition and monopoly

The argument really boils down to a question of performance. Does the monopolist behave against the community interest or does it achieve levels of efficiency beyond the capacity of small firms operating in highly competitive markets? There is no clear answer. As the extreme cases of monopoly are fairly rare in practice, examination is usually made of markets which approach monopoly conditions.

If the demand curve faced by the monopolist shifts, this will alter the marginal revenue curve and consequently the profit-maximising output and price. However, we cannot assume that the demand curve will simply move outwards parallel to the old one. It is possible that its slope may change, becoming steeper or less steep. Consequently, while normally we would expect an increase in demand at all prices to lead to an increase in monopoly price (assuming costs remained unchanged), we cannot be absolutely sure of this. Try experimenting with differently sloped average revenue curves. Remember that the marginal revenue must bisect (cut into two equal halves) the horizontal distance between the average revenue curve and the revenue (vertical) axis. You will find that there are changes that could produce a reduction in the profit-maximising price!

X-Inefficiency

The problem with the preceding arguments is that they assume that monopolists are efficient. The evidence is that large organisations, not just large firms with considerable market power, are inefficient when compared with smaller organisations and firms in competitive market situations. For example, large government departments and government-owned firms are notoriously inefficient. The UK National Health Service (NHS) is the third largest single organisation in the world (based on its number of staff). While the NHS is wonderful when you are in need of medical attention, many studies show that it is measurably inefficient and cost ineffective in comparison with both public and private health care providers in many other countries.

© ABE and RRC

The concept of *X-inefficiency* is used to explain the economic inefficiency of large organisations. At its simplest, X-inefficiency is a measure of the excess cost of production of a unit of output of a good or service by an organisation over the cost of producing the same output in the most efficient available organisation. Take an industry with two firms producing the same type and quality of good. Assume the two firms are of different sizes but there are no economies of scale. One firm has a unit cost of production for the good of £3 per unit, while the other firm has a unit cost of production for the same good of £4. The second firm is X-inefficient in comparison with the first firm. Its degree of X-inefficiency is 30 per cent. X-inefficiency in all types of organisations is ultimately the result of managerial failure.

The lack of the drive provided by the profit motive, and the threat of bankruptcy and closure for failure to keep costs down, means that bureaucratic organisations tend to be larger than necessary with far too many employees. They are also resistant to change, and tend to defend old, established or traditional ways of operation and prevent innovation, especially when such innovation would mean reducing the number of staff. The main reason for this inefficiency is the lack of an incentive in terms of a reward structure for workers to be efficient in carrying out their jobs. Workers are paid regardless of their individual work effort, usually simply on the basis of their hours at work. The absence of monetary reward or clear promotion prospects for working harder and/or longer than other workers means that most staff will behave in the same way, and follow human nature by taking things easy. Likewise if there is no reward for innovation in the way work is done or changing how departments are organised to reduce cost and increase output, there is likely to be an absence of change.

The problem is made worse by another feature of bureaucratic organisational structures in relation to the reward structure for managers. In many bureaucratic organisations, managers' pay and promotion prospects are directly proportional to the number of staff they have working for them. This means that mangers who increase efficiency and can deliver the same or more output with fewer staff damage their own pay and promotion prospects! The incentive structure is perverse, and rewards inefficient managers who can add to their department size and budget by demanding more and more workers to deliver the same or less output. Thus bureaucracies tend to be both cost inefficient for a given state of technology, and prevent or slow down technological innovation. The entire economy of the former Soviet Union was organised as a giant state bureaucracy and, not surprisingly, eventually collapsed because it was unable to match the efficiency and innovation that is a distinguishing feature of more market-orientated economies. It is difficult, if not impossible, to think of any modern consumer good or industry that originated in the former Soviet Union or China. Personal computers, mobile phones, the Internet and most consumer electronic goods all originated from the competitive environment in market economies and not from large bureaucratic organisations. It is no surprise that the economic transformation and success of China in global markets is a consequence of the reform programme introduced in the country in the late 1980s. In this process individuals were encouraged to start their own businesses and many state bureaucratic firms were broken up and privatised and encouraged to compete with each other in return for profit.

The concept of X-inefficiency is very important when evaluating the case for and against monopoly. Most arguments in defence of monopoly are based on the economies of scale in production that very large firms may experience, and the capacity of these firms to innovate resulting from their superior ability to fund and undertake research and development. But large firms are subject to the failings of large bureaucratic organisations. That is, the economies of scale that large firms (especially monopoly firms) are supposed to reap assume that they do not suffer from X-inefficiency. If increasing the size of a firm significantly leads to a reduction in unit costs of 25 per cent through technical and marketing economies of scale, but managerial slack resulting from bureaucratic complexity leads to a 30 per cent increase in its costs, the larger firm is less cost efficient not more cost efficient than smaller firms. Studies of efficiency in research and development (R & D) activity, and the sources of innovation in both processes and products, also show that large organisations suffer from X-

© ABE and RRC

inefficiency in undertaking R & D and are not the main source of process innovation in modern economies.

The advantages of monopoly are:

- lower prices than in competitive markets due to economies of scale
- larger expenditure than competitive firms on R & D
- more innovation due to large expenditure on R & D
- high level of investment expenditure because of large profits.

The disadvantages of monopoly are:

- higher prices than in competitive markets due to persistence of excess profit
- cost reducing advantage of scale economies outweighed by cost increases due to X-inefficiency
- wasteful expenditure on R & D and low productivity of R & D expenditure due to X-inefficiency
- no incentive to innovate because of high monopoly profit and absence of competition from other firms
- no incentive to investment in new production process and products because of existing high monopoly profit and absence of competition from other firms
- lack of customer focus – limited choice and poor product quality due to lack of competition.

© ABE and RRC

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. List the key assumptions of the economic model of a perfectly competitive market structure.
2. Why is a perfectly competitive market regarded as the ideal form of market structure?
3. How has the growth of the Internet affected competition in markets? Is eBay an example of perfect competition?
4. Explain the key characteristics of a monopoly industry. Can you identify any real world examples of a monopoly firm?
5. Using an appropriate diagram, outline the model of monopoly.
6. Compare the predictions, including equilibrium price, profit and deadweight loss, of the monopoly model of market structure with those of the model of a perfectly competitive market structure.
7. What is X-inefficiency? Why is it found in bureaucracies as well as large firms? Can you identify examples of X-inefficiency in any organisations with which you are familiar?

© ABE and RRC

© ABE and RRC

Study Unit 8

Market Structures and Competition: Monopolistic Competition and Oligopoly

Contents ***Page***

A. Monopolistic Competition **146**
Main Features 146
General Model 146
Comment 147

B. Oligopoly **148**
Price Competition 148
Price Stickiness 148
Kinked Demand Curve 149
Limitations of the Kinked Demand Curve Model 151
Price Leadership 152
Collusive Behaviour 153

C. Profit, Competition, Monopoly, Oligopoly and Alternative Objectives for the Firm **154**

© ABE and RRC

Objectives

The aim of this unit is to: explain the kinked demand curve model of oligopoly and the model of monopolistic competition; discuss the idea of collusion and identify the factors that affect the stability of a collusive arrangement; compare the predictions of these models with those of monopoly and competition.

When you have completed this study unit you will be able to:

- discuss the general characteristics of an oligopoly industry and identify the characteristic similarities and differences between oligopoly models and the models of perfect competition and monopoly
- identify, using the appropriate diagram, the characteristics of the kinked demand curve model of oligopoly
- identify the equilibrium price, output and profit in the kinked demand curve model
- explain why the kinked demand curve model predicts price stability and discuss the limitations of this model
- identify, using the appropriate diagram, the characteristics of the model of monopolistic competition
- identify the equilibrium price, output and profit in the model of monopolistic competition in the short and the long run
- discuss the meaning of collusion in the context of an oligopoly, examine the factors that aid or hamper the ability of firms to collude and discuss the implications of these findings for policy makers concerned with maximising social welfare
- discuss the price, output and welfare implications of oligopoly models relative to the models of monopoly and perfect competition.

A. MONOPOLISTIC COMPETITION

Main Features

Monopolistic competition still retains many of the features of perfect competition – unrestricted entry to and exit from the market, good (but not perfect) communication and transport conditions, motivation by economic considerations only, and the perception by buyers that the products of the various firms are good substitutes for each other.

It is in this last point that monopolistic competition differs from perfect competition. Although the products are considered to be good substitutes, they are not homogeneous. Buyers *do* express preference for one seller's product as opposed to another's.

Sellers seek to increase this preference by differentiating their product through branding (giving it distinguishing features) and especially by advertising. The greater the degree of preference they can establish, the stronger the brand loyalty and the greater the freedom gained by the supplier from the need to follow the market price for that class of product. Success brings an increased degree of market power and a reduction in price elasticity of demand.

General Model

However in the general model of monopolistic competition, we assume that the individual firm is not able to achieve a high degree of price inelasticity, so that the demand curve for the individual product has only a fairly gentle slope: there is still a high degree of substitutability between competing brands. This prevents the individual firm from making

© ABE and RRC

monopoly profits. It is still closely governed by the market price for the class of product. The result is shown in Figure 8.1.

In outline the features of this model are:

- There is no abnormal or monopoly profit: average cost equals price/average revenue at Op and, as for perfect competition and monopoly, it includes an element of normal profit.
- At the profit-maximising output of Oq, average cost is still falling to its minimum at Oc, where average cost is equal to marginal cost – the output level where the rising marginal cost curve cuts the bottom of the average cost curve.
- Price (at Op) is above marginal cost (Om) at the profit-maximising output Oq.

Price is thus higher and output lower than would be the case if price were to be equal to marginal cost, as in perfect competition. The lack of monopoly profit is the result of competition and the ability of firms to enter and leave the market.

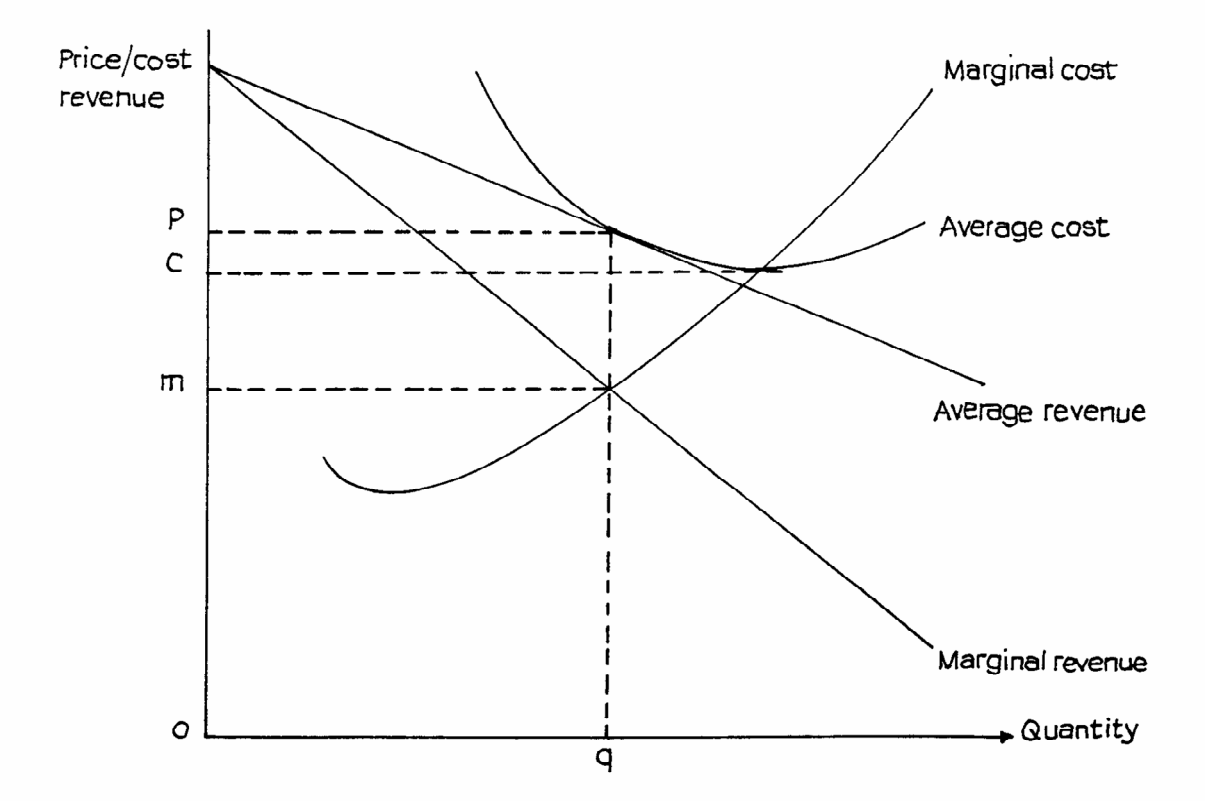


Figure 8.1: Monopolistic competition

Comment

It can be argued that this market structure is not really in the best interests of either consumers or business firms, for the following reasons:

- Price is higher and output lower than would be the case with perfect competition.
- The firm is not making the best use of its resources, since average cost is still falling at output Oq, as we saw.
- Profits are confined to the normal minimum required to keep firms in the market – the amount included in our definition of costs for the purposes of these market models. They cannot achieve the profits needed for investment and research or the high output levels necessary for economies of scale.

© ABE and RRC

That said, it is also argued that consumers are prepared to accept these additional prices and costs in return for the benefits they receive through greater choice of product – the ability to choose between competing brands and competing suppliers. This competition may also lead to improvements in product quality and design as well as services to the consumer.

We can expect firms operating in such market conditions to seek to increase their monopoly power and make their product-demand curves less elastic. They will do this by brand advertising, by securing favourable treatment from distribution organisations or through technical improvements in their products. They may be able to keep an advantage by securing patent protection or keeping processes secret from their competitors.

B. OLIGOPOLY

Oligopoly is the market structure where supply is controlled by a few firms which are large in relation to the market size. Very often the firms are also large by any standards, and are likely to be oligopolists in several markets. (For example, Unilever is a very large company which supplies major brands of many grocery products, including Marmite, Flora, Hellman's and PG Tips and washing products including Surf and Persil.)

Oligopoly is now commonly found in the advanced industrial countries and a great deal of attention is paid to it. However there is no single model which can be held to apply under all circumstances.

Price Competition

One influence that is thought to be important is the extent to which the products are in price competition with each other. If there is little price competition and if consumers are not thought to choose brands on the basis of comparative price (i.e. if cross elasticity of demand is low) then each oligopolist has a high degree of monopoly control over the demand for his own product.

This will of course depend chiefly upon whether the products are regarded by consumers as homogeneous or whether they consider each brand to be distinct and different. You might think it is unlikely that consumers will find much to choose between, say, various brands of plain, salted crisps. Cross elasticity of demand between the brands is thus likely to be high when the crisps are on sale in similar distribution outlets. If there are price differences, customers will choose according to price.

In these circumstances, suppliers may seek to operate in different sections of the market, e.g. through different supermarket chains or in hotels and pubs rather than retailers. They may also seek to differentiate their products through such devices as flavour or by developing novelty shapes or other related products. You may be familiar with various products which have been developed by the four major firms in this market.

A full study of oligopoly is likely to embrace problems of prices and non-price competition, and even the question of how far firms may collude together to limit the extent of competition between established firms and to protect themselves against possible newcomers to the market.

Price Stickiness

Efforts have been made to produce models based on traditional assumptions of profit maximisation. One such model seeks to explain the observed tendency that the prices of some goods in oligopolistic markets remain steady in spite of fluctuations in the prices of basic commodities. This "stickiness" is apparent in more normal, less inflationary times. For

© ABE and RRC

example, the price of bars of chocolate in some markets remains constant in spite of frequent movements in the prices of the basic materials required for chocolate manufacture.

This particular feature of an oligopolistic market for a product still regarded as fairly homogeneous (in spite of brand advertising) has given rise to the model known as the *kinked demand curve*.

Kinked Demand Curve

Suppose the current and "sticky" price of a product is £1 per unit. This is the price that customers have come to expect. If one oligopolist supplier tries to increase the price, rival producers will be reluctant to follow. They will keep their prices the same and gain market share at the expense of the price raiser. However if the oligopolist reduces the price, the other suppliers are obliged to reduce their prices also to prevent his encroaching on their market share.

Thus there is a kink around the price of £1 in the demand (unit price or average revenue) curve faced by the individual oligopolist. At higher prices the curve is more elastic, due to the loss of market share, than at lower prices, where all market shares stay the same. You can see the general shape of such a kinked curve in Figure 8.2.

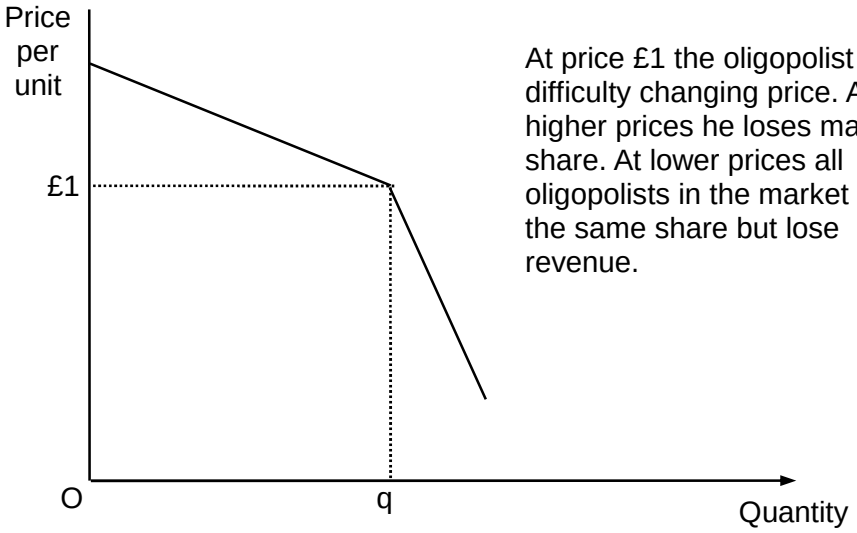


Figure 8.2: Kinked curve

Now consider possible revenues resulting from this condition, in Table 8.1.

© ABE and RRC

Price per unit	**Quantity**	**Total revenue**	**Marginal revenue** (Change in TR)
£	*units per time period*	*£*	*pence*
1.40	0	0.00	
			130
1.30	10	13.00	
			110
1.20	20	24.00	
			90
1.10	30	33.00	
			70
1.00	40	40.00	
			60 or 20
			0
0.80	50	40.00	
			–40
0.60	60	36.00	
			–80
0.40	70	28.00	
			–120
0.20	80	16.00	
			–160
0.00	90	0.00	

Table 8.1: Possible revenues

The kink in the average revenue curve, shown in Figure 8.3, occurs at the price of £1 and the quantity level of 40 units. At prices above £1, demand falls off at the rate of ten units for each 10p rise in price. At prices below £1 however, demand falls by only five units for each 10p rise in price, i.e. the unit price has to fall 20p to enable the oligopolist to gain a quantity increase of ten units.

The change in the slope of the average revenue (price) curve results in a similar change in the slope of the marginal revenue curve and you can see that there are two possible marginal revenues at the quantity level of 40 units. The higher (60p) results from the continuation downwards of the upper part of the curve, whilst the lower (20p) results from the upward continuation of the lower part of the curve. This is clearer on the graph but you should be able to work out the same results from the table. Remember the marginal revenue levels in the table belong to the midpoints of the quantity changes. The lower curve is changing at the rate of 40p for each ten units; the upper curve is changing at the rate of 20p for each ten units.

© ABE and RRC

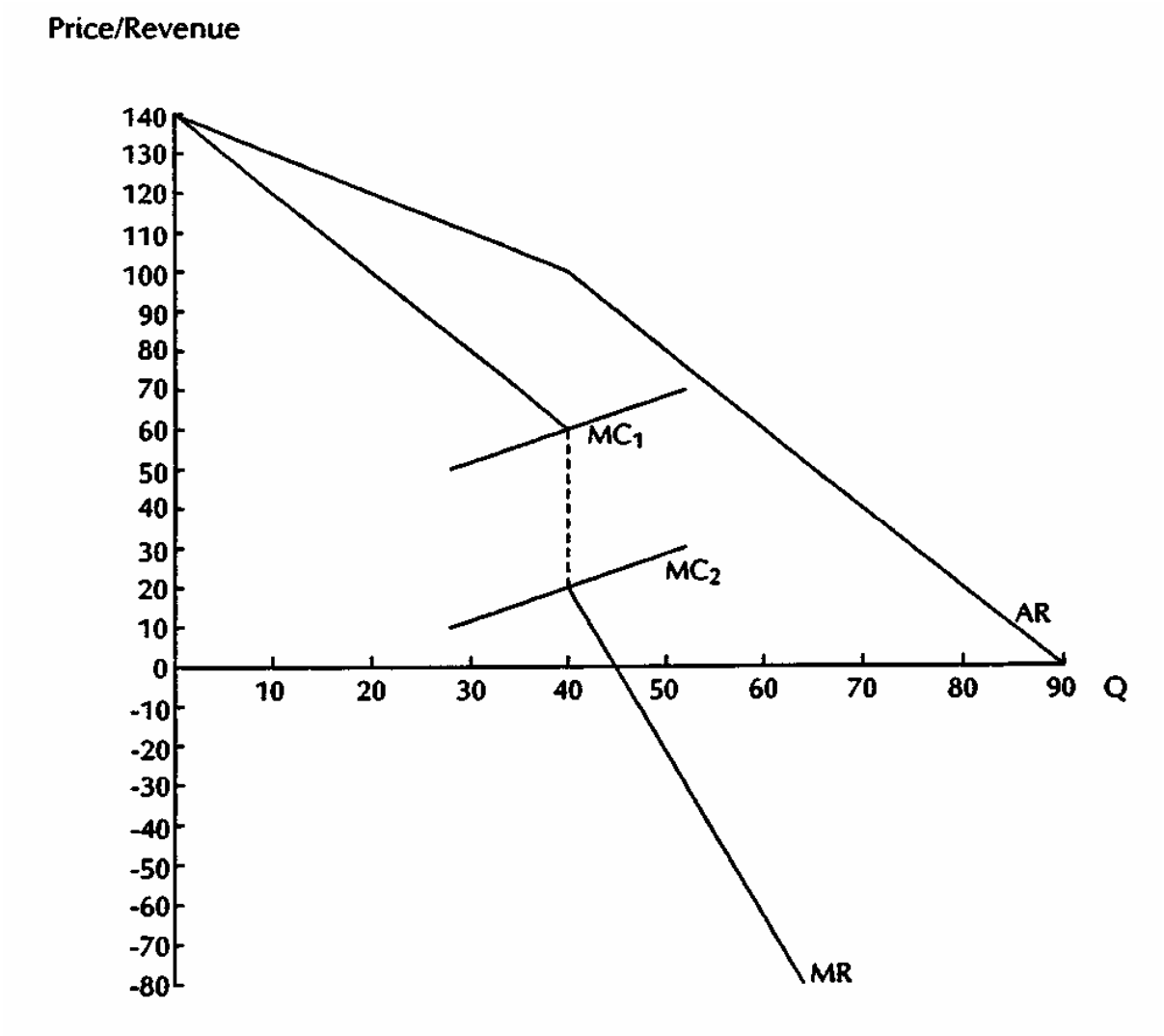


Figure 8.3: Quantity level at which profits are maximised

Limitations of the Kinked Demand Curve Model

The implication of this model is that short-term fluctuations of variable and hence marginal costs will not lead the profit-maximising oligopolist to change his price or output. You can see in Figure 8.3 that the quantity level at which profits are maximised (i.e. where MC_1 and MC_2 equals MR) is 45, at which level the market clearing price is 100p. Marginal cost can fluctuate anywhere between MC_1 and MC_2 without altering the profit maximising position.

Remember however, that this model depends on an assumption of profit-maximising behaviour for the oligopolist and a high degree of substitution between products. This produces the reactions from competing oligopolists that we have described (i.e. refusal to follow a price increase but matching a price reduction). It is not a general model of oligopoly and does not tell us how the "sticky" price is arrived at in the first place. There are too many behavioural assumptions for the model to be entirely satisfactory.

The model does not hold up during periods of severe price inflation, when we would expect firms to follow their rivals' price rises but not any price reductions which they will not expect to be maintained because of rising costs. Nor does it hold when there is a dominant firm acting as a price leader in the market.

© ABE and RRC

Price Leadership

Another tendency observed in some oligopolistic market situations is for the few firms in the market to follow the price movements of one firm, the *price leader*. Such leaders can be:

- The least-cost firm, which can oblige competitors with higher costs to follow its prices, even though they cannot maximise their own profits at the levels it sets.
- A firm which is typical of others in the market and which becomes a barometer of market conditions. If this firm feels that a price change is necessary, then it is probable that others will feel the same.
- The largest and the dominant firm in the market. The most common model of this situation assumes that this firm, because of its size and the economies of scale it can achieve, is able to achieve lower costs than the others. The lower its costs compared with the other firms' costs the greater will be its market share and, consequently, its dominance in the market. This model is illustrated in Figure 8.4.

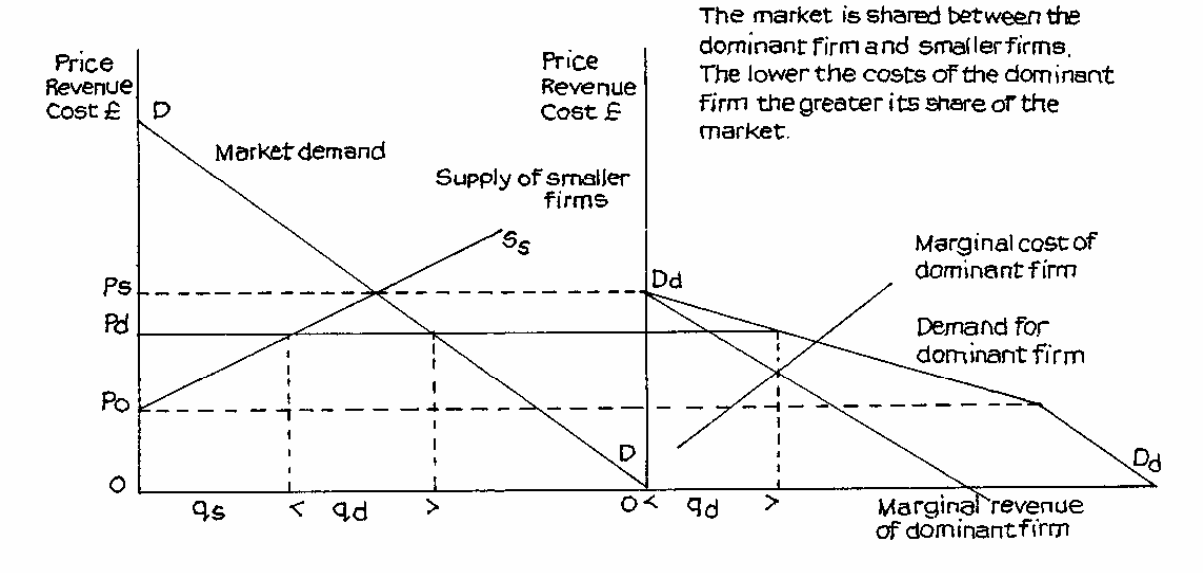


Figure 8.4: Price leadership model

The market is shared between the dominant firm and smaller firms. The lower the costs of the dominant firm the greater its share of the market.

The dominant firm model makes the following assumptions:

- The dominant firm is aware of the total market demand curve and the cost conditions, and hence the supply curve, for the smaller firms in the market.
- The objective of the dominant firm is to maximise profits.

In Figure 8.4 the demand curve DD is the demand curve for the market and S_sS_s is the supply curve for the smaller firms. At price P_o these firms are unwilling to supply to the market; it is their minimum price. At price P_s the smaller firms are able and willing to supply the full market demand at that price.

This knowledge allows the dominant firm to estimate its own demand curve, which is made up of market demand at each price less the amount which the smaller firms are able to supply. Thus the demand for the dominant firm's product is nil at price P_s but it is the same as market demand at prices P_o and below. Between these two prices the dominant firm is able to supply the balance between market demand and supply from the smaller firms.

© ABE and RRC

On the assumption of profit maximisation the dominant firm will wish to supply quantity q_d, which is the quantity at which its marginal cost is equal to its marginal revenue. At this quantity level the dominant firm's market clearing and profit maximising price is P_d. If it charges this price the other firms will have to follow, and market demand at this price is shared on the basis of q_d to the dominant firm and q_s to the smaller firms.

Notice that if you raise the dominant firm's marginal cost curve then you will reduce q_d and increase q_s. However, if you lower this curve you will increase the market share going to the dominant firm, which is thus able to maintain its dominance as long as it is able to keep its costs lower than those of the smaller firm. We may assume it is able to achieve this through economies of scale, a higher level of technical knowledge and managerial skill, and by its superior power to secure low prices in the factor markets.

Collusive Behaviour

Another distinguishing feature of oligopolistic market situations is collusion between firms in the industry. Although such behaviour, which includes *price fixing* (agreements to fix a common price), is illegal in many countries, the nature of oligopolistic market situations lends itself to collusive behaviour and agreements. Competition reduces prices and profits, which is why it is beneficial for consumers and the success of economies, but it makes life hard for the managers of companies and their owners who would prefer higher profits. In perfect competition the very large number of firms in the market makes it difficult for firms to get together and fix the market in their own interest. Oligopoly is different: because of the small number of firms, each one knows the others it is competing against. More importantly, each knows that if it changes its price, or any of the non-price features of its marketing, it will have an effect on the other firms' markets share and they will take action to restore their position. That is, oligopoly market situations involve *interdependence* between the behaviour of firms. Equally, the small number of firms in the market means that the owners/managers can easily arrange to meet and agree that if they stopped competing, reduced their outputs and set a common price, then they would all make more profit and have a quieter life! Recognising the independent nature of their price and output decisions, and the danger of a price war resulting from each firm trying to increase its market share/profits, leads firms in oligopolistic markets to collude and act as if they were one firm with monopoly power.

Such behaviour is more common than you might think: it often involves firms in different countries because many global markets, such as cement, steel and air cargo transport, are oligopolistic in nature. In the EU, where such collusive agreements are illegal, the Competition Commission has been successful in prosecuting firms which have fixed the price of glass, cement, plasterboards and vitamins. The US government has achieved a lot of success in fining firms for entering into collusive agreements.

Competition authorities try to prevent or break up collusive agreements between firms, to protect consumer interests against the monopoly exploitation such collusion is intended to achieve. Fortunately for consumers such collusive behaviour also contains the seeds of its own destruction, although it may take several years for the seeds to bear fruit, and consumers still lose out during this period. The instability of collusion in oligopoly and the reasons why collusion agreements break down include:

- The incentive for each member of a price-fixing and/or market sharing cartel agreement between firms to cheat on the other members. Once the cartel has set an agreed price, each firm will gain more sales and profit if it secretly cuts its own price below the agreed price. This will be done on the assumption that the other firms in the cartel obey the rules and keep their price at the agreed fixed level. Since every firm will reason in this way, each firm in the cartel has an incentive to secretly lower its price and/or try to sell more than its allocated share in another firm's market. The result of this individually rational behaviour by each firm is that they collectively destroy the price fixing and/or market sharing agreement!

© ABE and RRC

- Firms in the cartel are reluctant to share full information about their true costs, prices, sales and profit, or they give false information. This can lead to disagreements between members and lack of confidence that other members are sticking to the rules of the cartel. In turn this can lead to members responding to real or imagined rule breaking by other members by breaking the rules themselves.
- Firms in an oligopolistic market situation recognise that their price and output decisions are *interdependent.* The significant implication of this is that the normal relationships between price changes, and the consequent changes in sales and sales revenue, depend not just upon the elasticity of the firms demand curve. These relationships also depend upon how other firms respond to a firm in the market changing its price, as shown in the kinked demand curve model. This interdependence creates uncertainty for firms that have to determine their production and pricing decisions on the basis of *game theory*. The decision making is of the form: "*If I increase my price tomorrow by 10 per cent what will be the consequences for the other firms in the industry? How will they react? Will I still gain if they only decide to respond by increasing their prices by 5 per cent? What if my main competitor responds by reducing rather than matching my price increase?*" Each firm is in a *game situation*: think about the card players in a game of poker for a similar example. In such a situation it is highly likely that at some point one firm will make a decision to change its price and output, based on its assumption about the response of the other firms, and *get it wrong*. In this situation the market is unstable. A price war is a likely consequence, even when firms have a collusive agreement, if at least one firm to the agreement thinks that it can come out the winner in such a situation.
- Another reason for the instability of collusive agreements exists when such agreements are illegal. There is the incentive for one member to avoid legal prosecution, and a very large fine, by obtaining immunity from prosecution by being the first to spill the beans to the competition authority about the existence and details of the cartel. This is known as "whistle-blowing".

C. PROFIT, COMPETITION, MONOPOLY, OLIGOPOLY AND ALTERNATIVE OBJECTIVES FOR THE FIRM

In the discussions of perfect competition and monopoly, we noted that whereas under perfect competition long-term survival depended on the firm maximising its profits, whether or not this was its conscious objective, under monopoly the firm could survive without actually maximising profits. As long as it made a satisfactory profit it was able to pursue other objectives. We now develop this point more fully.

Any firm which possesses a substantial degree of market power as a producer and which is large in relation to the total size of the market in which it operates, will have a product demand curve which is downward sloping. If the firm is successful, it is also likely to be able to make profits above the minimum needed to keep it in the market. Its position may therefore be represented by a model similar to that usually used for monopoly as in Figure 8.5.

This model assumes that the firm does not practise price discrimination, so that its product demand curve is also its average revenue curve. Assuming that its market power allows it to make profits above the minimum, there will be a substantial range of output levels and prices between which it can make profits. This, in Figure 8.5, is the range between output level A (price P_A) which is the lower break-even point where the falling average cost just equals average revenue, and output level C (price P_C) which is the higher break-even point where the rising average cost just equals average revenue.

© ABE and RRC

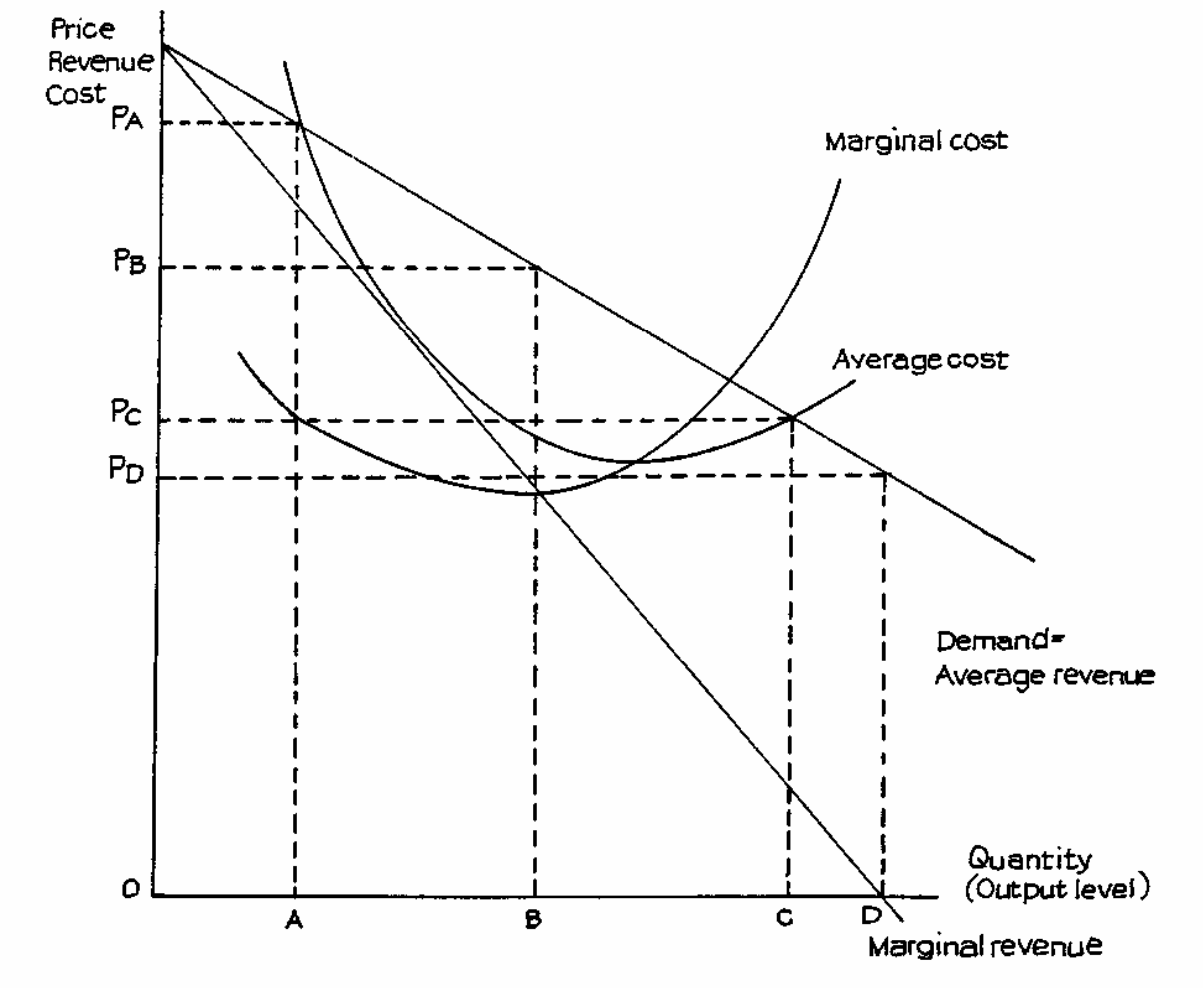


Figure 8.5: Oligopoly/monopoly model

The firm in this situation can pursue objectives other than profit maximisation as long as it operates within this profit range, but, as the model suggests, the range can be very wide.

A number of alternative theories of the firm have been developed and each of these is based on different assumptions about firms' behaviour. For convenience we can identify two broad groups of theories – those that replace profit maximisation by an assumption that firms seek to maximise something else, and those that abandon any idea of maximisation in the belief that firms seek to pursue several objectives at the same time and cannot therefore hope to optimise any one. Before looking at these alternative theories, which may have much more relevance for monopoly and oligopoly firm behaviour than for firms in competitive markets, it must be clearly understood that *no firm has a future unless it can cover its costs*. That is, all firms need profit to survive in the longer term. The assumption that all firms seek to maximise their profit is made to enable the development of models of firm behaviour. This assumption is simply the extreme limit of what all firms must do in reality if they want to survive. What the alternative theories do is provide additional rather than alternative insights into how firms might behave in practice provided they are profitable in the long-term.

(a) Alternative Maximising Theories

Baumol, an American economist, has suggested that firms seek to maximise *revenue*, subject to making a minimum profit defined as that level of profit needed to retain the support of the firm's shareholders and the financial markets. In Figure 8.5 the revenue-maximising output level is at D, where marginal revenue is O (at the top of

the total revenue curve). However in this model quantity D lies beyond the second break-even point of C, so the firm could not reach D without suffering a loss. If it were to try to maximise revenue subject to achieving minimum profit, it would have to produce at an output level somewhere between B and C and charge a price between P_A and P_B.

A British economist, Marris, has argued that firms seek to maximise their *rate of growth* (expansion), subject to preserving their share values at a level where the firm can hope to be reasonably safe from the fear of being taken over. If the firm grows too fast, its profit rate tends to fall and this depresses the share value and brings the risk of takeover. Too slow a rate of growth is also likely to bring the firm to the notice of take-over raiders, so the firm has to balance the desire for growth with the need to maintain profits.

There are similarities in the Baumol and Marris theories. Both agree that the firm's objectives are really established by its professional managers, who are free to control the firm as long as they keep the shareholders satisfied with their dividends and the financial markets satisfied with their profits. Profit remains important – no one doubts that in a market economy – but it is not *maximised* to the exclusion of other aims that meet managerial ambitions. Managers like to operate in large firms because size brings prestige, high salaries and a range of other benefits, so these are pursued, to some extent at the expense of the profits belonging to shareholders. In the Baumol theory, revenue is seen largely as a way of measuring growth. The Marris argument is slightly more complex and stresses growth more directly.

Another American economist, Williamson, developed another kind of maximisation, but quite cleverly combined this with the idea that the firm pursued several objectives at the same time. Again agreeing with the idea that managers were the real controllers of the firm, Williamson argued that they sought to maximise *managerial utility*. This utility was a combination of the pursuit of profit, growth (measured by the number of people employed), and managerial perks (all the various expenses, benefits, etc. that movement up the business managerial ladder tends to bring).

(b) Satisficing Theories

The rather ugly word "satisficing" has been coined to express the idea that firms pursue several different objectives at once. Whereas no one objective can be achieved to complete satisfaction, the firm aims to pursue each to a degree of tolerable semi-satisfaction, i.e. it "satisfices" without fully satisfying. The idea was first given clear expression by the American economist, Simon, in an influential book, *Administrative Behaviour*. Simon argued that in practice, firms could not, even if they wished, hope to maximise anything. Rather, they reacted to problems as they arose, and aimed to keep all those involved in the firm reasonably satisfied so that the firm could continue to exist.

Following the reasoning of Simon, this idea was developed into a more formal *Behavioural Theory of the Firm* by two more American economists, Cyert and March (in a book with that title). In this theory the firm is seen as a coalition between shareholders, managers and customers, all of whose support is needed to hold the coalition together. To achieve this support, the firm has to pursue multiple objectives, such as profit, sales growth, market share and products to satisfy customers as well as the needs of production managers, but no one objective can be pursued to the exclusion of the others. The firm has to develop a set of behavioural principles to enable it to hold the coalition together and guide managerial decision-making.

Various other attempts have been made to explain business behaviour, but there is no general agreement as to whether the traditional assumption of profit maximisation should be abandoned and, if so, what should replace it. The alternative theories sometimes seem to

© ABE and RRC

describe actual business behaviour more realistically, especially in relation to large oligopolists. Firms do pursue growth, often at the expense of profits, takeover battles are commonplace and the salaries and prestige of top business managers appear to bear little relationship to the profitability of the companies they manage.

On the other hand, an economic theory of the firm should be concerned not only with how firms actually do behave but also how they *should* behave, if the economic goals of technical and allocative efficiency are to be achieved. Unfortunately, the alternative theories appear to suggest that if firms operate as they predict, they are likely to be less efficient in the full economic sense than if they pursue profit maximisation – the desire to make the largest achievable profit consistent with market conditions. One thing that has to be remembered always is that profit maximisation does *not* mean making very large and antisocial profits, but simply the largest profit possible under prevailing market conditions. Profit maximisation under perfect competition suggests lower profits than satisficing behaviour in an oligopolist market. A market economy appears to operate more efficiently when firms seek to maximise profit. Consequently, most economists continue to work with profit-maximising models, whilst fully recognising that firms do frequently depart from profit-maximising behaviour in practice.

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. Outline the main features of the model of monopolistic competition.
2. How does the equilibrium of a firm in a monopolistically competitive market differ from that of the firm in a situation of perfect competition or that of monopoly?
3. Identify some examples of a market structure that resemble that of the economic model of monopolistic competition.
4. Explain the characteristics of an oligopoly industry. Identify some examples of oligopoly market situations.
5. Using appropriate diagrams, explain the kinked demand curve.
6. List some of the forms of collusion undertaken by firms in an oligopoly industry.
7. Explain why collusive arrangements between firms in an oligopoly tend not to be sustainable in the longer run.

© ABE and RRC

© ABE and RRC

Study Unit 9

The National Economy

Contents ***Page***

A. National Product and its Measurement 160
Flows of Production and Money 160
Flow of Production and Consumption 161
The Consumption Function 163
Modifications to the Basic Flow 163
National Product, Income and Expenditure 164
National Income – Treatment of Taxes and Subsidies 165

B. National Product 166
Avoiding Double Counting – Value Added 166
Gross Domestic Product 167
Trends in Domestic Product 168

C. National Expenditure 169
Calculation of GDP 169
Gross and Net National Product 170

D. National Income 170

E. Equality of Measures 172

F. Use and Limitations of National Income Data 173
Reasons for Introduction of National Accounts 173
Helping to Solve Economic Problems 173
Making Comparisons 173
Limited Accuracy 174
Value to the Community 174
Changing Money Values 174

H. National Product and Living Standards 176

© ABE and RRC

Objectives

The aim of this unit is to evaluate national income as a measure of societal well-being and derive it through its various methods of measurement.

When you have completed this study unit you will be able to:

- compare and contrast expenditure, income and output measures of national income
- explain the distinction between gross *domestic* income and gross *national* product
- demonstrate an understanding of nominal and real measures of national income
- identify the different treatment of taxes, subsidies and transfer payments in the national income accounts
- explain how national income per capita is measured and the limitations of relying upon this measure
- recognise other, non-economic, aspects of well-being
- explain how broader indices of well-being work, for example, the Human Development Index.

A. NATIONAL PRODUCT AND ITS MEASUREMENT

Flows of Production and Money

In this study unit we start to examine the national economy as a whole. We see this in terms of one large market, in which total or aggregate demand from the whole of the community is satisfied by total production. We are thus concerned with totals or aggregates in this part of the course. When we have gained an understanding of the national system, we can begin to see its interrelationship with the wider international economy.

We are concerned chiefly with modern industrial economies – or with agricultural economies organised on an industrial basis (e.g. states such as Denmark or the Republic of Ireland). Some of the important assumptions which we shall be making will be valid for these economies but would have less relevance for subsistence agrarian economies, organised around self-sufficient local communities, or for completely state-regulated socialist economies.

Data on aggregate economic activity in the UK is published each year in the *United Kingdom National Accounts* (the publication which is also called the *Blue Book*). One of the key sets of data in the accounts is that for gross domestic product (GDP for short). In the *UK National Accounts* GDP is defined as "the sum of all economic activity taking place in UK territory". Economic activity is explained as follows:

> "In its widest sense it could cover all activities resulting in the production of goods and services and so include some activities which are very difficult to measure. For example, estimates of smuggling of alcoholic drink and tobacco products, and the output, expenditure and income directly generated by that activity, have been included since the 2001 edition of the Blue Book."
> (*United Kingdom National Accounts – The Blue Book 2006*, page 8)

Economic activity or production generates output and:

> "this economic production may be defined as activity carried out under the control of an institutional unit that uses inputs of labour or capital and goods and services to produce outputs of other goods and services. These activities range from agriculture and manufacturing through service producing activities (for example financial services and hotels and catering) to the provision of health, education, public administration and

> defence: they are all activities where an output is owned and produced by an institutional unit, for which payment or other compensation has to be made to enable a change of ownership to take place."
> (*United Kingdom National Accounts – The Blue Book 2006*, page 8)

Flow of Production and Consumption

The *national economic concepts* we use assume that:

- Production and consumption are separate – production being organised by business or government organisations, and consumption being decided by individuals, families and households. The family is thus seen as purely a consumption and social unit, and not as a production/consumption/social unit, as it would be in an agrarian (farming) economy.
- Most of the goods and services produced are exchanged through a market system, with households paying money to buy products, and firms paying money for the use of production factors.
- A proportion of production is organised by the state and its agencies, and paid for by revenue raised by the state from the community.

This system can be illustrated in the form of two circular flow diagrams, see Figure 9.1. One shows the flow of goods and services – the productive activities of production factors (Figure 9.1(a)), while the other (Figure 9.1(b)) shows the counter-flow of money which oils the really important flow of production and consumption. Notice that for simplicity, we use the terms "firms" for production organisations, and "households" for the individuals and families who consume what is produced. These diagrams assume that the total volume of production is immediately and totally consumed, i.e. there is nothing to enlarge or diminish this continuous circular flow.

Notice that firms are seen as hiring the production factors, which are owned by households, which then supply the labour, capital and land employed in production, and purchase the goods and services produced.

© ABE and RRC

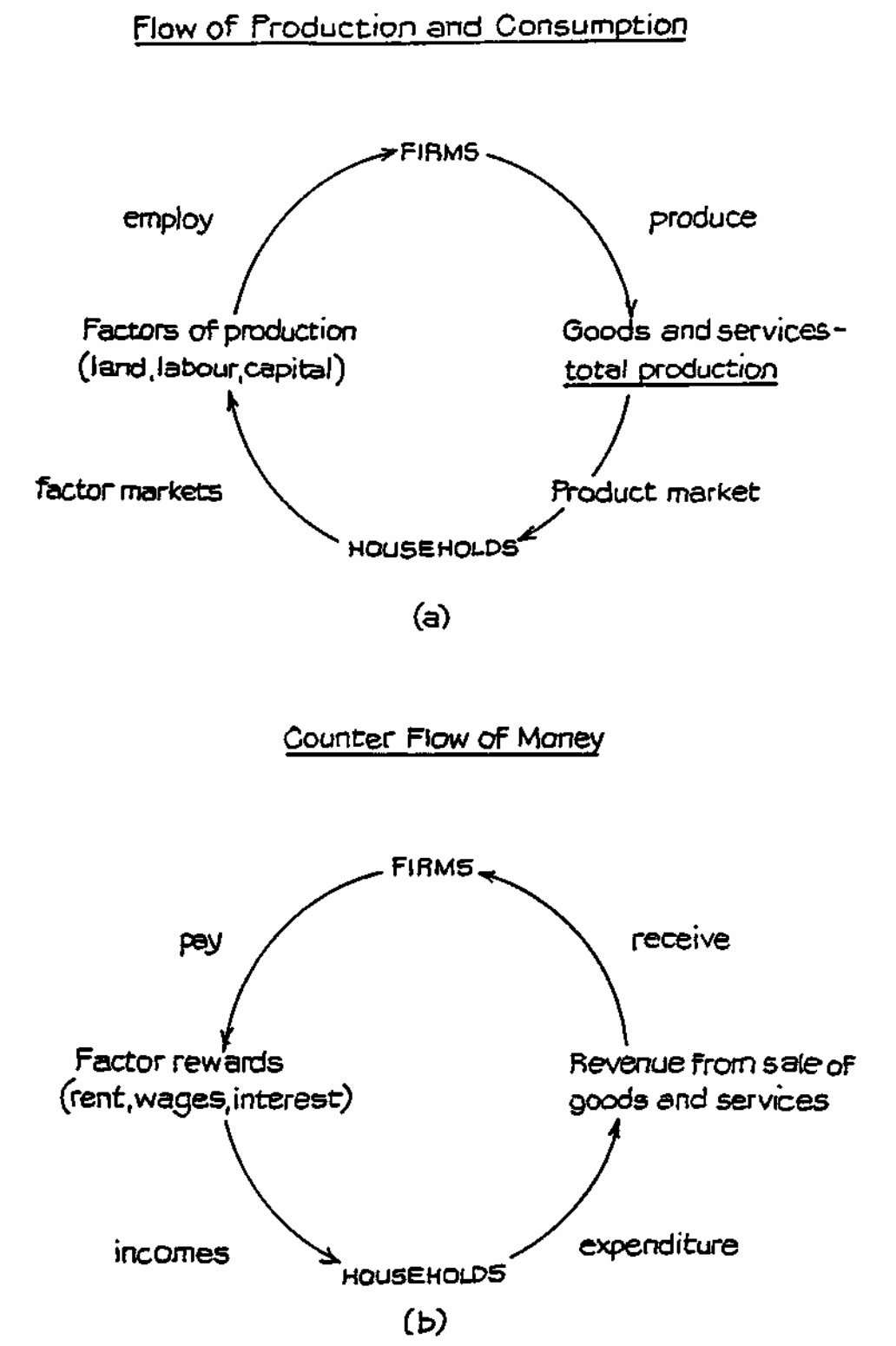


Figure 9.1(a): Flow of goods and services and Figure 9(b): Flow of money

© ABE and RRC

The Consumption Function

If, for simplicity, we imagine an economy where there is no foreign trade, no taxation and no government spending, then we can say that total income is either spent (consumed) or not spent (not consumed). If we then define savings as income that is not spent or consumed, then we can make the proposition that income (Y) is either consumed (C) or saved (S), i.e. that:

$$Y = C + S$$

Given this proposition and retaining our simplified model of the economy, we can then see that any increase in income is apportioned between consumption and saving. The amount of any increase in income which is consumed is often referred to as the *marginal propensity to consume*. It may also form the basis for an equation which helps us to determine the level of consumption for any given level of national income. For example, we may say that:

$$C = 300 + 0.75Y$$

This is then termed the *consumption function*. The term "function" will be familiar to you from your study of mathematics and quantitative methods.

Given this function, i.e. the direct relationship between total consumption and total income, we can calculate values for C for any level of Y. For instance, if:

$$Y = 1{,}000, \text{ then } C = 300 + 0.75 \times 1{,}000 = 1{,}050$$

At this level, people are trying to consume more (1,050) than their total income (1,000) and will have to use up past savings or borrow from another country. At the income level (Y) of 4,000:

$$C = 300 + 0.75 \times 4{,}000 = 3{,}300$$

This means that savings will equal 700, i.e. 4,000 − 3,300.

In this example, the 300 is a constant; it is the minimum amount of consumption required by the community, whatever the level of income. Total consumption is made up of this minimum plus a proportion of total income. The greater the marginal propensity to consume, the higher will be the proportion of total income that is consumed at any given income level. If the marginal propensity to consume remains the same at all income levels, then this will also be the proportion of Y that is consumed in the equation.

Modifications to the Basic Flow

We must now modify some of the assumptions made in the basic circular flow concept. The main modifications we need to make are to take into account the following factors:

(a) Not all the income received by households is immediately spent on goods and services; some income is *saved*.

(b) Another part of total income of households is not actually spent on goods and services but handed over to government authorities as taxation, either taken directly from income or indirectly when certain goods and services are purchased. At this stage, all forms of taxation are considered together. We shall examine forms of taxation later.

(c) Yet another portion of income is spent on goods and services produced by other national economies, i.e. it is spent on *imports* from other countries.

(d) Firms enter the general flow as buyers of goods and services, such as factories, machines and research, in their efforts to increase their capacity to produce. We call this *investment* or *capital accumulation*.

(e) The government must be seen as a separate force which produces goods and services on behalf of the community as a whole – e.g. it builds roads, schools and

© ABE and RRC

hospitals, and it provides forces to maintain law and order and defence against external aggression. We can combine all these activities under the heading *government expenditure*.

(f) Firms supply other countries with *exports* of their products. Trade is a two-way process.

We can regard modifications (a) to (c) as leakages from the main flow of economic activity, because they reduce the purchasing power of total incomes. We can regard (d) to (f) as injections into the flow, because they increase total purchasing power and demand. This concept of leaks from and injections into the main flow is illustrated in Figure 9.2.

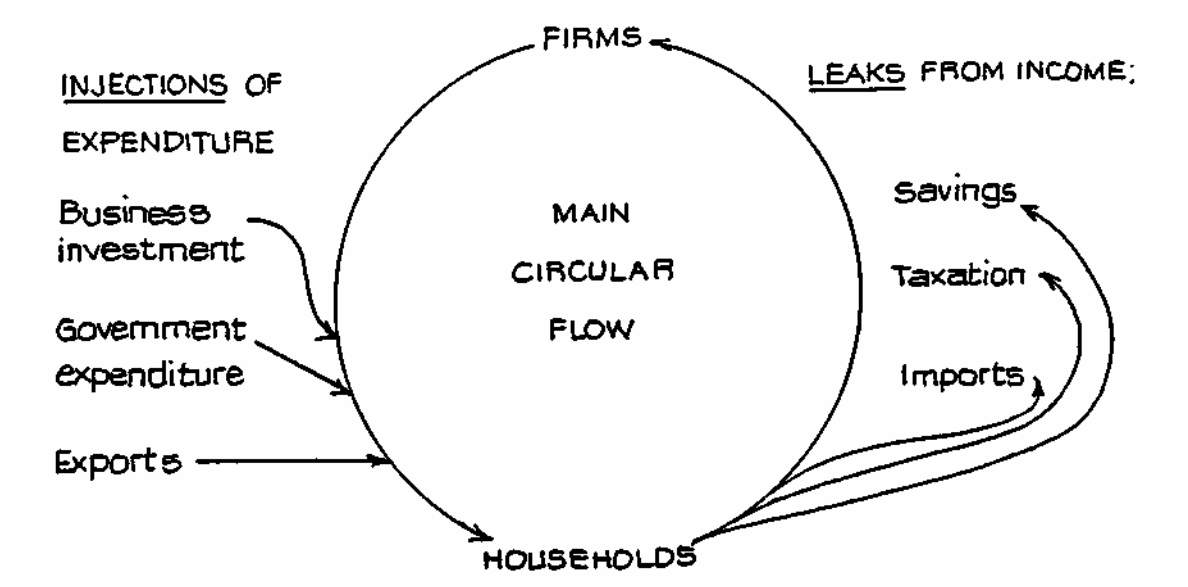


Figure 9.2: Leaks and injections into the main flow

National Product, Income and Expenditure

This total flow of economic activity, modified by injections and leaks, can be given the general term *national product*. This is the term used chiefly today, and it serves to emphasise that it is the total *production* of goods and services that is the really important matter. This is the total flow as seen in our first illustration (Figure 9.1(a)). The counter-flow of money in the second diagram (Figure 9.1(b)) can be seen as both the total *income* of households and as the total *expenditure* of households.

Notice that these three – total product, total income and total expenditure – are all really describing the same essential flow. They can be regarded as equal – provided that the total amount of leakages from income (savings, taxes and imports) is equal to the total amount of injections of expenditure (from investment, government spending and exports).

At the moment, we shall assume that this equality does exist and that total production equals total income equals total expenditure. Thus, if we use P to denote total product, Y to denote total income, and E to denote total expenditure, we can say that:

$$P = Y = E$$

We therefore need to examine each of these aspects of the flow more carefully.

© ABE and RRC

National Income – Treatment of Taxes and Subsidies

It is useful here to examine more closely the treatment of taxes and subsidies in the national income summary accounts calculated from incomes and from expenditure.

The national account actually show two versions of gross domestic product based on expenditure. One, at *market prices*, takes no account of expenditure taxes or subsidies paid to producers. This measure shows the totals of spending at the prices actually paid "in the market". The other measure of GDP is calculated by deducting the total value of expenditure taxes and other indirect taxes and adding back the total of subsidies paid to producers. This measure is commonly referred to as the *"factor cost"* measure as it shows the "*true*" cost of production of output, since indirect taxes are not a true cost of production despite the fact that they appear as part of the cost when the goods and services are purchased. Likewise subsidies reduce the prices paid for goods and services below their true cost of production.

However the *UK National Accounts* are now constructed in accordance with the 1995 European System of Accounts (ESA95) and the term (but not the concept) "factor cost" is no longer used. The term "*basic prices*" is now used in place of factor cost.

The difference between factor cost and basic prices involves the distinction between those indirect taxes that are levied on each unit of output, and those indirect taxes, such as the tax on vehicles, which are levied on producers (the production process). This is not a difference you need worry about: if you prefer, you can continue to use the term "factor cost" instead of "basic prices" to refer to national output net of indirect taxes and plus subsidies. However, when looking at the *UK National Accounts* you will have to remember that the term "factor cost" is rarely used today.

A national product based on basic prices is the one normally used. It is considered to be the fairer reflection of true expenditure on goods and services. After all, total expenditure includes government spending on final consumption, and much of this is paid for from expenditure taxes. If we value GDP at market prices, then we are in effect including expenditure taxes twice – once when they are paid by the consumer, and once when they are used to pay for goods and services by the various government bodies. Similar adjustments need to be made to take account of subsidies. These are payments made by government to producers and have the effect of reducing market prices. To obtain the true cost of goods and services any subsidies need to be added back. An explanation of the meaning of *basic prices* is given in the *Blue Book*.

> "These prices are the preferred method of valuing output in the accounts. They reflect the amount received by the producer for a unit of goods or services, minus any taxes payable, and plus any subsidy receivable on that unit as a consequence of production or sale (i.e. the cost of production including subsidies). As a result the only taxes included in the price will be taxes on the output process – for example business rates and vehicle excise duty – which are not specifically levied on the production of a unit of output. Basic prices exclude any transport charges invoiced separately by the producer."
> (*United Kingdom National Accounts – The Blue Book 2006*, page 9.)

The *Blue Book* also explains the meaning of *purchasers' or market prices*:

> "These are the prices paid by the purchaser and include transport costs, trade margins and taxes (unless the taxes are deductible by the purchaser)."
> (*United Kingdom National Accounts – The Blue Book 2006*, page 9.)

The treatment of direct (mostly income and profits) taxes appears on the surface to be rather different, but the effect is the same – i.e. to ensure that total incomes are a fair reflection of the incomes actually earned in the course of producing the national product.

© ABE and RRC

Income and profits taxes are *not* deducted from employment incomes, nor are they deducted from the trading profits of companies and the trading surpluses of government-owned bodies.

The gross incomes, profits and surpluses are the true incomes actually paid by the production organisations.

On the other hand, no account is taken in the summary totals of incomes from pensions, unemployment benefits or other state welfare payments. These incomes are *not* received in return for a contribution to production. They are *transfer payments* – being transfers from the income of a contributor to the production process to someone who is a "non-producer". (No moral judgment is intended here. The non-producer may have been a valuable past producer, or he or she may become a valuable future producer. Our concern is to arrive at a true valuation of production *in the year of account*.)

The accounts do of course include the incomes of those in the employment of state organisations, even though their incomes may have been paid for out of income taxes. This does not matter – the incomes of state employees are earned in return for their work which is included as part of total production, and the process is no different, in principle, from any other use of income to provide an income to another *in return for goods or services*. If I use part of my income to pay for my daughter's dancing lessons, then those payments are included again in the accounts as part of the dancing teacher's income. If part of my income is taken from me to pay the salary of a teacher in my daughter's comprehensive school, then again, these payments are included in the national accounts. The only difference is that the state directs what I shall pay towards teaching in the school, whereas I choose whether or not to pay for the dancing lessons. In each case, the payments are made in return for services which contribute towards the production of the national product. What is not included as a further income is the payment made out of my taxes towards the unemployment benefit paid to my unemployed nephew. His income is not earned in the course of producing anything, and it is ignored, as though it were a voluntary contribution from me to him.

B. NATIONAL PRODUCT

Avoiding Double Counting – Value Added

The national product is the sum of the values of all the goods and services produced by a community within a recognised time period – normally a calendar year. However, we cannot simply add up the values of all goods and services produced by all business organisations in the country. If we did this, we would be counting some things more than once. For example: a set of screws may be made by firm A, sold to firm B which makes timing equipment, which in turn is sold to firm C – a motor-vehicle assembler. The completed vehicle is then sold to firm D, a motor distributor.

The final price of the vehicle includes the cost of the screws but, if we added up the total value of the products sold by firms A, B, C and D, we would find that we had counted the screws four times.

One possibility might be to add up only the value of the products sold by the final distribution firm, but this might not give us a very accurate result. This is because our motor distributor does not always know whether they are selling to a householder, or to a small business firm which will use the vehicle for business purposes and include its cost in the value of the goods or services it produces. There would also be considerable problems of allowing for goods imported and exported.

The solution actually adopted is to count the "value added" to inputs by all firms producing outputs. This is now much easier than in the past, because of the introduction of *value added*

© ABE and RRC

tax (VAT). All firms paying the tax are in effect also reporting their value added to the taxation authorities. In very simple terms, the value added by each firm is the difference between the revenue it obtains from selling its product and the cost of all goods and services purchased from other firms. In this way, the screws of our original example are counted only in the value added of firm A. They are excluded from the totals obtained from firms B, C and D. Notice that value added includes the cost of labour employed by each firm in adding value to the inputs it purchases.

We shall go on to show how public sector spending contributes to the gross national product. However, there is a reservation that should be made when we consider the public sector. This concerns what are often called *transfer payments*. For example consider what happens when a person receives unemployment benefit or some similar social security benefit. This is not a payment made in return for work performed or services provided. It is a transfer to the unemployed person through taxation from the income earned by people in employment. If we counted the unemployment benefit into the national product in addition to the full income of those who in effect are making the transfer, then we would be double-counting the amount. Incomes are counted as part of national product only if they are earned by some contribution to economic activity, e.g. by employment or by making capital available to government or business. Payments received by way of transfer through taxation are not included in the total – though, of course, they have to be taken into account when we examine how the total national product or income is distributed.

A similar transfer payment within the private sector takes place when parents give pocket money to their children. The income has been earned by the parent and is simply transferred to the child. Total national accounts thus do not include children's pocket money! Of course, the transfer payments taking place through the public sector are much larger, and it is important that we understand why they should be excluded from the final totals.

Gross Domestic Product

The figures published in the *Blue Book* show total product figures classified by categories of industry and service. The following table is adapted from the *Blue Book 2006* and shows the figures for 2004.

© ABE and RRC

Industry sector	2004 *£ million*
Agricultural, forestry and fishing	10,323
Mining and quarrying including oil and gas extraction	21,876
Manufacturing	147,469
Electricity, gas and water supply	17,103
Construction	64,747
Distribution and hotels	160,594
Transport, storage and communication	79,279
Financial intermediation (net), real estate, renting and business activities	294,350
Public administration and defence	55,280
Education, health and social work	137,603
Other services	55,543
Gross value added (GDP): all industries at basic prices	1,044,165
(Basic prices is almost the same as the old factor cost method of measurement)	

Table 9.1: Gross domestic product by industry: gross value added at basic prices

Total domestic output of products represents the gross value added by all the economic activity of the community, measured from the output of business and government organisations. This figure is termed *gross domestic product* (GDP). The basis on which this figure is valued does not include indirect taxes and government subsidies, so that it is valued at "basic prices", i.e. at the cost of the factor inputs, not at the prices paid by final consumers.

Trends in Domestic Product

The largest item in the domestic product in 2004 was that relating to financial and business services, a sector which accounted for over 28 per cent of the domestic product, outstripping manufacturing (under 14 per cent) which for years had been the largest sector. The decline in manufacturing's share of total product has been continuing for many years as services of all kinds have assumed an increasing importance. This is a trend that is common to all the old industrial countries of North America and Western Europe. It reflects both rising living standards in these countries, where people spend an increasing proportion of incomes on services instead of goods, and changes in the pattern of world production. (Look at the goods manufactured in the Pacific Rim countries of Japan and South East Asia.)

If you compare the figures in Table 9.1 to those of previous years, you will also notice the rise in the proportion of product accounted for by education, health and social work. This has occurred for a number of reasons: changes in technology affecting the work performed and equipment used by these services; the age structure of the population as the rising numbers of older people put more pressure on the health services; and changes in economic and social conditions, with the expansion of education to cope with the demands of the modern technology-based society and of social work to cope with the casualties of that society.

© ABE and RRC

The relative growth of services at the apparent expense of manufacturing does not mean that manufacturing is no longer important to economies such as that of Britain. It is still extremely important, not only because the financial and business services need a strong manufacturing base for their own development, but also because it still provides a very large share of the wealth of the community. Manufacturing has of course changed. It is no longer made up of simple "metal bashing", but is based on complex, computer-aided processes often involving very high levels of technology. The borderline between the new manufacturing processes and services is often rather vague. Assembling a computer is clearly a manufacturing process. However designing the software and systems that control the computer and all the other equipment in the factory depends on the services of teams of designers and programmers, who would not think of themselves as working in manufacturing.

Further developments may also be slightly exaggerating the trend away from manufacturing towards services. The old-style manufacturing firm employed many groups of workers in-house, such as caterers and designers and those performing other commercial service activities. Today these jobs are more likely to be carried out under contract by specialised services firms, but they are still actually performed for the manufacturing firm and its workers. These statistics, like all others, need to be interpreted with some caution and against a background awareness of what is actually happening on the ground.

C. NATIONAL EXPENDITURE

Calculation of GDP

The main items of total expenditure were identified earlier as the main flow of household consumption plus the injections of business investment, government spending and export demand. The concept is reflected in the *Blue Book* totals which, in the 2006 edition, identified the national product by category of expenditure in 2004, as follows:

Category of expenditure	***£ million, current prices***
Consumption expenditure	761,484
Central government consumption	152,325
Local government consumption	98,383
Total gross capital formation	199,310
Total domestic expenditure at current prices	1,211,502

Table 9.2: National product by category of expenditure for 2004

Consumers' expenditure is the same as the household expenditure already explained. Total central and local government spending is shown exclusive of capital investment. For example, it includes the running costs of the Health Service but not the cost of building hospitals. This capital investment or formation is combined with private sector investment to produce the fourth item in the table, "total gross capital formation". The sum of these categories of expenditure is *total domestic expenditure.*

This figure is not the same as domestic product calculated from industrial and government output, because of the effect of imports and exports. Consumer and other spending will include spending on goods and services produced in other countries (imports), but will not

© ABE and RRC

include the value of goods and services sold to other countries. However, when we add on a figure of £298,694 million for exports, and deduct £333,669 million for imports, we obtain the total for *gross domestic product calculated from expenditure* of £1,176,527 million at market prices. This is not the same as the figure calculated from value added (production) by industrial sector given earlier because that figure was at basic prices (factor cost if you like).

The basis of the valuation of production is at basic prices (factor cost), because the effect of taxes and subsidies on expenditure has been removed. If we add to the figure of £1,044,165 million for *gross value added at basic prices* given earlier (Table 9.1) the *Blue Book* figures for indirect taxes, and subtract the figure for subsidies, we will arrive at *gross domestic product at market prices*. This used to be referred to as the "factor cost adjustment": in 2004 £139,642 million taxes less £7,280 million in subsidies, a total factor cost adjustment of £132,362 million. This is the amount by which gross domestic product, measured at current market prices, would be overvalued by the effects of taxation and subsidy. Factor cost gives the true value of the production factors used to produce the total product. Thus in 2004:

Gross value added at basic prices	=	£1,044,165 million
Adding back taxes on expenditure	=	£139,642 million
Subtracting subsidies	=	£7,280 million
Gives us gross domestic product at market price	=	£1,176,527 million.

Gross and Net National Product

The *Blue Book* makes two further adjustments to the GDP total. These are given next.

(a) An allowance for "net property income from abroad": earnings of British organisations operating in other countries less the amount earned in the UK by foreign-owned organisations. Actually the relevant figures have to be adjusted for compensation of UK employees received from abroad and paid abroad, i.e. migrant workers remitting part of their earnings back home. They also have to be adjusted for taxes paid to the rest of the world and subsidies paid overseas. In 2004 there was a net inflow of this income of £26,525 m and when this is added to gross domestic product at market prices it gives us a total of £1,202,075 m. This is known as the *gross national income at market prices.*

(b) An allowance for "capital consumption": the using up of capital investments made in past years (e.g. the deterioration of roads, factories, machines, computers, etc.). In 2004, this was estimated to total about £128,427 m. Thus, when this figure is deducted from the gross domestic product of £1,176,527 m, there remains a total for *net domestic product* of £1,048,100.

In a similar way, if we deduct the figure for capital consumption from *gross national income at market prices* we obtain *net national income at market prices*. If net national income at market prices is converted to basic prices, adjusted for indirect taxes and subsidies, we arrive at the figure for *net national product at basic prices* which is the measure termed *national income* in the national income accounts.

In practice, the figure most commonly used for international comparisons etc. is that for gross national product – largely because the capital consumption figure has to be estimated and different countries use different methods of estimation.

D. NATIONAL INCOME

We noted earlier that total factor incomes suffered leaks from savings, taxes and import spending before they were transformed into expenditure. The main *Blue Book* totals do not in fact show these items directly, although they can be calculated from figures published in

© ABE and RRC

the book. Instead, they show the gross national product by category of income. The totals are of gross incomes, so they include taxation, savings and money which will be spent on imports. The categories of income are as in Table 9.3.

Notice that these correspond broadly to the rewards to factors of production. Compensation of employees (income from employment) is the return to labour, although this may also include some return to business owners' capital in the case of the income of the self-employed. Gross operating surpluses of corporations correspond to profit of private companies and government organisations (including public corporations). These surpluses may be seen as the reward to capital.

The table also shows that the sum of all the incomes generated in an economy within a year are equal to the gross value added at factor cost of all the economic activity that takes place in the economy. The addition of taxes on products and production less subsidies, plus an adjustment for any statistical discrepancy between the production and income methods of measuring national output, gives us the figure for total GDP at market prices, shown in the final column.

As we are concerned with incomes earned within the country, we do not have to make any adjustments for imports and exports.

YEAR	Compensation of employees Total	Gross operating surplus of corporations[1] Total	Other income Total[2]	Gross value added at factor cost Total	Taxes on products and production less subsidies	Statistical discrepancy (income) Total	Gross domestic product at market prices Total*
1995	386035	174186	69372	629593	93487	0	723080
1996	403887	191345	76301	671533	97372	0	768905
1997	429967	200659	80449	711075	104806	0	815881
1998	466080	205944	81806	753830	111880	0	865710
1999	495793	207971	86723	790487	121458	0	911945
2000	532179	210488	87842	830509	128422	0	958931
2001	564194	211196	97352	872742	130555	0	1003297
2002	587396	235819	97468	920683	135110	0	1055793
2003	616893	257629	102494	977016	141229	0	1118245
2004	648717	280180	106183	1035080	149216	0	1184296
2005	686805	282320	114149	1083274	151618	-916	1233976
2006	723143	301093	121304	1145540	159377	-1344	1303573

The main components of income leading to gross domestic product at market prices.

Seasonally adjusted; £ million at current prices.

** Note that the figures given in the final column differ slightly from those given for GDP at market prices in the rest of this unit because they are based on revised data.*

1 Quarterly alignment adjustment included in this series.

2 Includes mixed income and the operating surplus of non-corporate sector less the adjustment for financial intermediation services indirectly measured (FISIM).

Source: ONS online statistics 2008

Table 9.3: UK national income categories of income 1995-2006

© ABE and RRC

E. EQUALITY OF MEASURES

Notice that the *Blue Book* – and countries other than the UK use similar calculations – takes care to emphasise the equality (or, more strictly, the identity) of the three measures by:

(a) ensuring that each is brought to the same total, where necessary by the device of a "statistical adjustment"; and

(b) labelling each set of summary accounts as "National or Domestic Product" – thus stressing that it is the *same flow of activity that is being measured*, whether by category of expenditure, category of income or class of industry.

This also emphasises that it is *real output*, i.e. the flow of *actual goods and services*, that is important, rather than the flow of money through income and expenditure patterns.

Thus, the national account supports the concept of national product and its circular flow. Remember that total gross incomes were distributed by households as: consumer expenditure, savings, taxation and spending on imports.

At the same time, total expenditure received additions (injections) from investment, government spending, and spending on exports by foreign countries. Bearing in mind that total income and total expenditure are different ways of looking at what is, essentially, the same flow, we can use symbols to state an equation. We have already used E for total expenditure and Y for total income. In addition to these, it is usual to make use of the following:

S = savings

I = investment

T = taxation

G = government spending

C = consumer expenditure

X = exports

M = imports.

Using these symbols, we can now say that:

$$Y = C + S + T + M$$

and

$$E = C + I + G + X$$

Remember that Y = E, so that:

$$C + S + T + M = C + I + G + X$$

Consumer spending (C) is common to both sides of this equation, so that we can expect the remaining elements of total income and total expenditure to preserve the equality:

$$S + T + M = I + G + X$$

This is a proposition which is of very great importance in our analysis of national product, and we shall be analysing its implications in some detail later.

© ABE and RRC

F. USE AND LIMITATIONS OF NATIONAL INCOME DATA

Reasons for Introduction of National Accounts

The detailed calculation and publication of annual national product figures is a practice with only a relatively short history. United Kingdom figures have been compiled regularly only since the early 1950s. If the nation managed to survive fairly successfully through the centuries before 1950 without national accounts, why do we attach so much importance to them today?

The answer is twofold. In the first place, the national product concept based on the circular flow of economic activity is relevant only to an industrial economy, and the UK could be called such only from around 1850 onwards. The realisation that the periodic economic problems arising out of industrial activity could not be measured and properly understood unless accurate figures were available, led eventually to acceptance by the government of its duty to prepare these figures.

The second part of the answer lies in the changed economic role of the government. After the Great Depression of the 1930s, there was a widespread belief that the government could and should seek to become involved in some degree of economic planning. If a government is to try to manage the national economy, it needs national accounts, just as much as business managers need business accounts for the firms they are seeking to control.

Helping to Solve Economic Problems

The existence of national accounting figures also helps us to understand how an economy actually works. Without precise figures, we can only guess at such issues as the influence of interest rates on savings or of income levels on consumption. When we have continuous records of interest rates, savings, incomes and consumption over a reasonable number of years, then we can produce evidence of cause and effect.

The more we know about the workings of a modern economy, the more hope there is that action can be taken to produce results that are beneficial to the community, and that solutions can be found for the great problems which beset industrial societies, such as mass unemployment and price inflation.

Making Comparisons

Accounting records make comparisons possible. We can find out whether the economy is operating more or less effectively than in the past, or more or less efficiently than the economies of other countries. As we shall see in the next section, care has to be taken in making comparisons but, without national accounting figures, no comparison is possible at all. For example, when we look at the UK experience over the last decade in the light of, say, the West German experience over the same period, we can see that there have been very different results arising from different policies and objectives.

One very practical use for national income figures is as a basis for a number of United Nations calculations. Member contributions to some UN institutions depend on their national product. National income and product figures are the starting point for many UN investigations designed to improve the economic and social performance of poorer countries. However, we have to accept that too much reliance should not be placed even on the best national accounts, and they should not be used, except with very great care, for purposes for which they were never intended.

© ABE and RRC

Limited Accuracy

It is clearly impossible to compile details of all the many economic activities in a modern community. The desire to evade taxes is one of many reasons why some activities remain firmly hidden from official eyes. The extent of the hidden (or black) economy in some countries is sometimes put as high as 20–50 per cent of the official economy! Business organisations come and go, and it is not easy to estimate the size of activity in new industries or the extent to which older activities may be declining. We have seen that the three measures of the British national product can be made to balance only with the help of a statistical adjustment. Considering the huge amounts involved the proportional differences that have to be reconciled are remarkably small. In countries able to devote fewer resources to statistical services the margin of error is likely to be rather greater.

Remember that we are dealing with large aggregates or total figures, and these can conceal very wide variations. For example, if on the basis of our accounts we say that the average income per head of the population is £x, we should not imagine that the majority of people will be earning that figure. Some will be earning much more and some much less. Some of the richest people in the world come from the poorest countries. For a developing country, any average is likely to be very misleading in view of the very great social, regional and other differences that exist.

Some countries may have an interest in ensuring that figures are not too accurate. A country hoping to obtain maximum help from, and make the smallest possible contribution to, United Nations institutions will wish to keep its national income figures as low as possible.

There is also the problem of comparing accounts when these are prepared in different national currencies. International figures are usually converted to United States dollars at official rates of exchange. Such official rates are often very different from the rates ruling in unofficial currency markets.

Value to the Community

So far, we have identified problems of calculation. Even if all the calculations and estimates were completely accurate, some important economic activities would not be included at all in the accounts. The most commonly-quoted example of a major omission is that of the contribution made to economic and social welfare by unpaid mothers, and others who perform services within the family. In the same way, official figures ignore unpaid voluntary activities within local communities and amateur sporting activities.

The way in which production, especially service production, is valued may cause further problems. Where goods and services are distributed through unregulated markets, we accept that market price is a fair method of arriving at their value. However where the state is the sole provider of a service and the sole employer of the factors used to produce that service, then we cannot be sure that the recorded value bears any relation to the value to the community – or to their value in another country where similar services are distributed through the market system.

Hospital charges in the USA, where there is a free market in health care, are higher than in the UK, where the National Health Service is the main supplier, and nurses earn more in the USA than in the UK. In Britain, charges in private commercial and language schools are higher than in the state-controlled colleges of further education. These differences make fair comparisons extremely difficult.

Changing Money Values

Any comparison or calculation is likely to rely on money as a measuring device. However, measuring any product with money is a bit like measuring a metre of cloth with an elastic rule. Money itself does not keep a constant value. Its value is eroded by price inflation. The rate at which prices increase (or sometimes decrease) differs greatly over time and from

© ABE and RRC

country to country. The rate of change in prices in a country can be measured using price indices, and in many countries various price indices are compiled for this purpose. These cannot be entirely accurate, and the longer the period over which comparisons are made, the less reliable the figures become.

In the *UK National Accounts* allowance for changes in the value of money is incorporated into the figures. This is done by a process of price adjustment referred to as the *"chained volume measurement method"*. The resultant figures are referred to as *"real values"* because they measure actual changes in output rather than changes resulting solely from changes in prices. This makes it possible to look through "the veil of money", and observe and compare "true" changes in output or income. Thus in seeking to establish the true extent to which economic progress is taking place in an economy over time it is necessary to use measures of real GDP or real national income. If the population of a country is also increasing it is necessary to express measures of real income or product on a per capita basis (real GDP per capita equals total real GDP/total population, and real national income per capita equals total real national income/total population).

Summary of National and Domestic Income and Product Relationships

You may find it helpful at this point to see in summary form how the different national accounting concepts and terms used in the *UK National Accounts* we have discussed are related.

GDP = gross domestic product (or income) at market prices

less primary income payable to non-residents

plus primary income received from the rest of the world

equals

GNI = gross national income at market prices

(this is equal to the sum of gross primary incomes received by resident institutional units and sectors of the economy)

and

real GDP (GDP converted from money value using the chained volume measurement method)

plus trading gain

equals

RGDI = real gross domestic income

plus primary real incomes received from the rest of the world

less real primary incomes payable abroad

equals

RGNI = real gross national income (converted from money value using the chained volume measurement method)

plus real current transfers from abroad

less real current transfers abroad

equals

RGNDI = real gross national disposable income

© ABE and RRC

The money and real values for the economy's measures of GDP, GDI, and GNDI are converted to their equivalent net values, NDP, NDI, and NNDI, by subtracting the estimate for capital consumption or depreciation.

For example, GDP less fixed capital consumption gives NDP. Because of the difficulty of calculating accurate measures of an economy's annual depreciation in its capital stock – its capital consumption – estimates of GDP are the most widely used measures of an economy's economic activity and the most reliable for comparisons between countries. For example:

GNI *minus* capital consumption *equals* NNI = *national income*

and

RGNI *minus* real capital consumption *equals* RNNI = *real national income.*

H. NATIONAL PRODUCT AND LIVING STANDARDS

All the points outlined in the previous section suggest that we should be very careful indeed if we use national product or national product per capita or per head (total national product divided by the number of people in the country) figures for the purposes of measuring living standards. We should take particular care when we make comparisons between countries with different economic and social systems, or attempt to measure changes over long periods of time.

Imagine an extreme case – an attempt to compare average living standards between 1888 and 2008. There was no radio, television, mobile phones, personal computers, portable music players such as iPods, or motor cars and aircraft in 1888! These are so fundamental to the pattern of life today that we cannot really even begin to make any sensible comparison. At best, we can only compare different aspects of life, e.g. working conditions, for particular groups of workers.

Moreover, when we talk about the standard of living, there are important aspects that cannot be measured in terms of economic activity. A person may have a higher real income if employed in 2008 than their father had in 1988, but if they are unemployed and have little prospect of employment, is their standard of living any higher? Opportunities for travel, for changing employment, freedom of speech and religion, freedom to walk the streets without fear of violent crime, arbitrary arrest or political coercion, all these are elements in the standard of living which are not included in any gross national product calculations. The matters of working hours and leisure time are also ignored. There is also the environment. Some countries attach great importance to protecting their environment and preventing pollution and other actions that degrade the physical environment. In other countries the environment may be ignored in both private and government decisions, and the physical environment may be so damaged and polluted that it damages people's health and reduces living standards. Standard measures of national income take no account of environmental damage and differences in the quality of the environment between countries. Some countries today, such as China and India, are achieving very high rates of real economic growth using conventional measures of national income, but at the expense of large scale damage to their physical environments (including their supplies of water). Material living standards measured by real GDP per capita can increase at the same time as the quality of life deteriorates and the former is the cause of the latter.

Economists are sometimes accused of placing too much weight on measures of quantity and on money values, and not taking sufficient notice of quality and the values that money cannot measure. Increasingly however, economists are recognising the limitations of the concepts and measures they use. As long as we bear these in mind, then we can make effective use of national accounts and recognise that these are an essential starting point for

© ABE and RRC

any study of national economy. We would not expect a set of company accounts to tell the full story of a large business enterprise but equally, if we wanted to examine the enterprise, we would be foolish not to include in that examination a very close scrutiny of the company accounts. In the same way, we find a great deal of invaluable information in the national accounts of a country.

Table 9.4 summarises the factors that need to be taken into account when using official measures of national income or GDP to compare changes in living standards over time in a country or between countries at the same time:

Income comparisons over time in a country	***Income comparisons between different countries at the same point in time***
Correct for changes in the level of prices over time – use real value measures by adjusting money values for inflation/deflation.	Compare like with like and use real value measures of GDP or national income.
Allow for changes in the size of the population – use real income per capita measures by dividing real GDP or real national income by total population. But need to recognise that real GDP per capita is an average measure and ignores how actual income levels per head are dependent on the distribution of income.	Compare like with like and adjust for differences in size of population by comparing real GDP or real national income on a per capita basis But need to recognise that real GDP per capita is an average measure and ignores how actual income levels per head are dependent on the distribution of income.
Allow for changes in the distribution of income over time in making conclusion based on changes in real GDP per capita	Recognise that differences in the distribution of income between countries affect conclusions based on a direct comparison of living standard measures such as real GDP per capita.
Allow for improvement in the quality of goods and services over time and the introduction of totally new goods and services.	Allow for differences in the quality of similar goods and services, and the availability of different goods and services, between countries at the same time.
Need to take account of changes in measures of the quality of life including health care, life expectancy, education and literacy, political freedom, press freedom, corruption, environmental pollution. For example, the Human Development Index.	Need to take account of differences in the quality of life between countries including health care, life expectancy, education and literacy, political freedom, press freedom, corruption, environmental pollution. For example, the Human Development Index.

Table 9.4: Key factors using official measures to compare changes in living standards

© ABE and RRC

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. What is meant by the circular flow of income? How does the circular flow of income in a closed economy differ from that in an open economy?
2. Explain the output, income and expenditure approaches to the measurement of gross domestic product (GDP).
3. Describe the main components of total expenditure or demand in an economy.
4. Explain how indirect taxes and subsidies are accounted for when we calculate an economy's GDP at basic prices (or factor cost) from the components of total final expenditure.
5. Explain the distinction between real and current price (nominal) measures of national output, product and income.
6. What is the term used to distinguish real from current price (nominal) measures of national output, product and income in the *UK National Accounts – Blue Book*?
7. Why are measures of national economic performance such as GDP or GNI not necessarily good guides to the standard of living or well-being in a country?

© ABE and RRC

Study Unit 10

Determination of National Product: The Keynesian Model of Income Determination and the Multiplier

Contents ***Page***

A. Changes in Consumption, Saving and Investment **180**
Equilibrium Conditions 180
Pressures Leading to Equilibrium 181
Pressures to Change Equilibrium 182

B. Government Spending and Taxation **184**

C. Changes in Equilibrium, the Multiplier and Investment Accelerator **185**
Equilibrium, Savings and Investment 185
Change in Investment and Change in National Income 186
The Investment Multiplier 188
More Realistic Multiplier 188
Change in the Marginal Propensity to Save and the Paradox of Thrift 189
The Investment Accelerator 190
The Business Cycle 191

D. The Role of the Government in Income Determination: the Government's Budget Position and Fiscal Policy **192**

© ABE and RRC

Objectives

The aim of this unit, in conjunction with Study Unit 11, is to explain the determination of the equilibrium levels of national income using the Keynesian macroeconomic model in a closed and open economy and demonstrate how this can be of use to businesses.

When you have completed this study unit *and Study Unit 11* you will be able to:

- interpret, graph, and solve simple numerical examples of the form $Y = C + I + G + (X - M)$
- explain how variations in the marginal propensity to save, consume, and import affects the closed and open economy multiplier
- compare and contrast inflationary and deflationary gaps using Keynesian cross diagrams
- discuss the components of government fiscal policy and explain how changes in these components affect the equilibrium level of national income
- make judgements about the factors that determine the effectiveness of fiscal policy
- explain the implications of fiscal policy for government borrowing (Public Sector Borrowing Requirement).

A. CHANGES IN CONSUMPTION, SAVING AND INVESTMENT

In this study unit we introduce the basic model of national income determination and the concept of the multiplier. These form the framework for analysing the process of determining the level of total or aggregate output in an economy, and the concept of macroeconomic equilibrium. The Keynesian model of national income determination and the concepts of the multiplier and macroeconomic equilibrium provide: the framework for understanding the means by which governments can use fiscal policy; the power of governments to tax and spend in the economy; and the power of governments to alter the levels of output and employment in the economy. This is such an important part of the syllabus, and a challenging one when studied for the first time, that the topic is studied in this study unit and in Study Unit 11. This means that the learning outcomes detailed in this unit can only be achieved fully after you have completed your study of both units.

Equilibrium Conditions

We should now remind ourselves of the conditions necessary for national product, income and expenditure to be in equilibrium. Remember the term "equilibrium" refers to a state of rest where there are no pressures acting to disturb and change the balance of forces. Earlier, we suggested that there would be equilibrium when total income was equal to total expenditure in the economy, and that this implied:

$$C + S + T + M = C + I + G + X$$

where:

C = personal or household consumption

S = savings

T = taxation

M = imports

I = business investment

© ABE and RRC

G = government spending

and X = exports.

If we remove C from each side of the equation, we are left with:

$S + T + M = I + G + X$

Putting this another way, we could say that total leaks or withdrawals from income equal total injections or additions to aggregate expenditure.

Equilibrium suggests that the state of rest remains for a period of time, so that we should take successive time periods into account. If, for simplicity, we use the symbols:

W = total withdrawals (S, T, M) and

J = total injections (I, G, X),

then, using the usual symbols t, t+1, t+2, etc. for successive time periods, we can say that a total national product in equilibrium implies that:

$W_t = J_{t+1} = W_{t+1} = J_{t+2}$, and so on.

Pressures Leading to Equilibrium

It seems reasonable to question why a national economy should achieve and maintain this form of equilibrium. If we examine the processes operating within the economy, we can see that there are strong pressures likely to produce such a state. For simplicity, we shall at this stage omit imports and exports from our analysis.

To begin with, we shall also omit taxation and government spending. We are now considering only *savings* and *investment*. However we shall reintroduce consumption.

Consider the graph shown in Figure 10.1. Expenditure intentions at the various national income levels are recorded in the curve C + I. Remember that we have reduced total spending to consumption and investment, for our present purposes.

Assuming that the scales of both axes are the same, then the 45° dotted line represents all points where total income just equals total expenditure. Remember too that when expenditure equals income, both are also equal to total output.

The graph illustrates that there is only one level of income where total income, output and expenditure are in fact equal – i.e. where national product is in equilibrium.

This is at the income level Oy_e, where the intentions curve intersects the dotted 45° line. However what happens if this equilibrium is disturbed?

(a) Lower National Income

Suppose national income is at the lower level Oy_1, where intentions are trying to achieve a higher level of spending than that possible from current total output.

At level Oy_1, the combined demand from households (C) and business firms (I) is higher than total output.

It cannot be satisfied at the current level of output. Some firms will have stocks of goods produced earlier, and they will be able to sell from these stocks. Others, finding that they have more customers than goods to sell, will ration sales by putting up the price or promising delivery at a future date. Actual consumption and investment will thus be lower than intended, as some would-be buyers are disappointed, but also money spending will be raised by the increased prices of goods.

© ABE and RRC

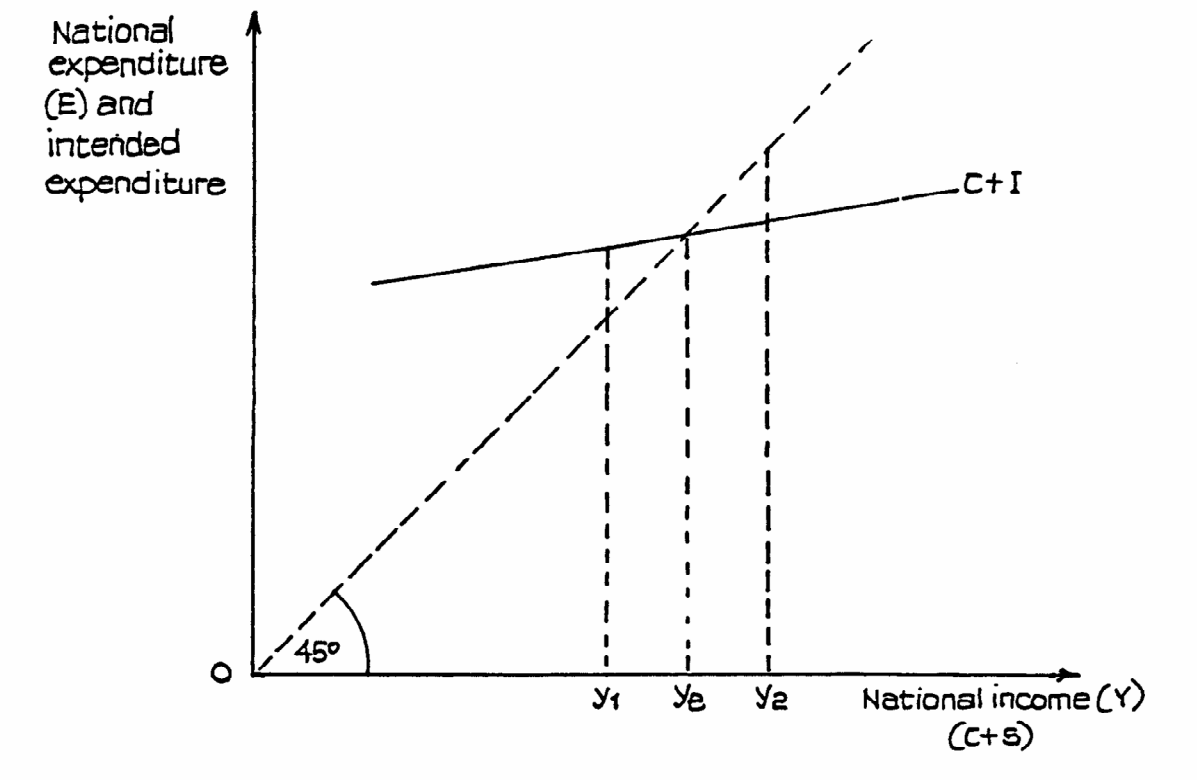


Figure 10.1: The national product in equilibrium

Increased money spending will feed into increased money incomes, and so the money value of national income will move up towards Oy_e. We can also expect that firms, facing high demand and good profits from rising sales, will seek to increase production. They will hire more labour and pay more wages in order to do this. This will tend to push up production towards Oy_e. There will be an upward pressure to achieve at least the money level of Oy_e, even if this still leaves many spending intentions unsatisfied.

(b) Higher National Income

We can apply this reasoning in reverse if national income happened to move out of equilibrium to the higher level Oy_2. Here, more is being produced than people want to buy. Warehouse stocks rise, and customers are not around to buy the goods and services on offer. Traders needing money to meet current expenses will cut prices to achieve sales. Firms, seeing stocks of unsold goods rise, will reduce production, lay off workers and cut overtime working. Incomes will fall through declining wages and falling business profits. There will be a movement downwards towards the equilibrium level Oy_e. Only at this level will there be no pressures for moving either up or down, because only here does total income equal total output equal total expenditure.

Pressures to Change Equilibrium

If we look again at Figure 10.1, we can see that this is only a stable equilibrium, lasting over successive time periods, if the curve of C + I remains unchanged. The higher we raise the C + I curve, the greater will be the level of Oy_e.

So although there are strong pressures to bring national income to equilibrium, there may also be forces operating to change the position of the C + I curve, and so change the equilibrium. In order to understand these forces, we need to examine more closely the decisions that lead to any given level of *desired* or intended household consumption and *desired* or intended business investment.

© ABE and RRC

(a) Influences on Aggregate Consumption and Saving

Remember that we have dropped imports and exports from our simplified model, and at the moment we are ignoring taxes and government spending. Therefore all income can be considered to be made up of consumption and saving. To emphasise this, we adopt a wide definition of saving, seeing it as *any income* (net of tax) not consumed. Thus for each unit of income:

$$C + S = I$$

or, $S = I - C$

Why then do people spend on consumption and why do they save? We can identify the following motives:

(i) They spend because they have income available for spending and perhaps because they expect future incomes to rise.

(ii) They spend because there is credit available.

(iii) They may also spend because they expect prices to rise and the cost of credit may be less than the amount of the expected price rise.

(iv) Pressure to spend may also come from advertising and the marketing efforts of firms wishing to maintain high levels of production and sales. Social attitudes may also encourage a high level of spending, especially in a period when the level of social security payments is high and money is losing its value and discouraging saving.

(v) On the other hand, saving may be encouraged and spending discouraged by falling incomes and rising unemployment, by controls on credit and the expansion of money, and by expectations that prices may fall.

(vi) People may also be forced to spend less and save more in order to pay off or reduce the burden of past debts after a period of high spending. This tendency was clearly evident during the early years of the 1990s after the spending and house purchase boom of the 1980s.

(vii) The depressed, low consumption years of the early 1990s also showed the importance of house purchase as a foundation for general household consumption. When house purchase and building activity is high and people are moving homes, they also spend on house furnishings, household equipment and so on. When there is little activity in the housing market all these associated household consumer durable markets are depressed. Employment and incomes fall in the affected industries and the economic depression deepens.

(viii) Savings may also be contractual, i.e. people undertake to save regular amounts out of income through schemes arranged with insurance companies, building societies, etc. The motives for contractual saving are to provide for retirement, for substantial future purchases, for precautionary motives, or simply because of social habit – the belief that saving is a moral duty.

Some of these motives correspond with the influences on the demand for products, which we identified in earlier study units. The general influences on total or aggregate spending and saving can change over time, so that the amount saved from any given volume of income can also change. Relationships between the amount consumed and saved and total incomes are examined later in this unit.

© ABE and RRC

(b) Business Production and Investment

Just as (leaving aside government spending, taxes, and foreign trade) we find that total income is either spent on consumption or saved, so we see total production as being sold either for consumption or for investment or capital accumulation. Here we have a slight problem: we cannot, in practice, distinguish between the purchase of new equipment to replace old and worn-out equipment, and that purchased to increase productive capacity. Moreover, some equipment may also be acquired simply to replace labour, with no significant increase in production planned or desired. Also, when we define investment in terms of production not sold for consumption, this includes stocks of goods.

So not all total investment could really be called "productive investment", able to increase the ability of business organisations to produce more. Yet, it is productive investment that really interests us. For simplicity, at this stage we shall assume that all or most investment does have a productive element (after all, most firms replace machines with better machines). This enables us to link the desire of firms to invest with their desire to produce more output. Thus, we can suggest that the main motive for investment is the belief of business firms that it will be in their interests to increase productive capacity. They are more likely to believe this if:

(i) current consumer demand is rising and expected to continue to rise

(ii) current profits are rising and expected to continue to rise

(iii) the cost of investment is falling and expected to continue to fall – the main element in this being the level of interest rates charged on borrowed finance.

Notice here that the influences on the level of investment are mostly not directly related to the level of current income. So for our purposes at this stage, we do not regard the level of investment as being dependent on income levels. This is in contrast to the level of saving which, provided other influences are constant, is directly related to the level of income.

Note that business firms, in making investment decisions, stress the importance of estimates of future revenues related to present costs and how these are affected by expectations of future demand levels and the costs of capital (linked to market rates of interest). You should remember that investment decisions involve making judgments about the future, about future markets and about future economic conditions and government policies. The future can never be forecast with accuracy, but the greater the degree of uncertainty about the future, the higher are the risks of business investment and the less the amount of investment undertaken. Political uncertainties and lack of confidence in the government can be as damaging to investment as market uncertainties; in practice the two are closely related.

B. GOVERNMENT SPENDING AND TAXATION

We now return to government spending and taxation, and seek to examine the relationship which exists between these. Of course taxation is the main source of government revenue, and if a government pursues a policy of a "balanced budget" (i.e. if it seeks to spend only what it earns through revenue), then the amount of spending must be governed by the amount of taxation received.

However, if a government does not believe that it must maintain this balanced budget, then the level of spending is released from the constraint of taxation and depends solely on policy decisions made by government ministers. We cannot therefore know what the influences on this spending are, unless we know the policy objectives of the government. Possible objectives and the economic ideas underlying different policies will be examined later.

© ABE and RRC

You may wonder how a government can escape from the constraint of its taxation revenues in deciding how much to spend. The answer lies in its power to borrow, and this power is itself a major influence on the economy. If the government borrows from the public directly, e.g. through national savings certificates and bonds, it will simply transfer income from the private to the public sector. If however it borrows from the banks, then it will be creating money. This is a difference that will have some significance for economic policies.

Taxation must come, either directly or indirectly, from income. It may come directly from taxes on private incomes and company profits, or indirectly through taxes on expenditure, such as value added tax. Since consumption expenditure levels depend on income levels, we can say that the total level of taxation is dependent on income.

C. CHANGES IN EQUILIBRIUM, THE MULTIPLIER AND INVESTMENT ACCELERATOR

Equilibrium, Savings and Investment

If we assume once more that we have an economy where the government has a balanced budget, so that taxation equals government spending, and imports just balance exports, then we can concentrate again on savings and investment. Under these conditions, national income will be in equilibrium when savings equal investment. This is illustrated in Figure 10.2. Another way to illustrate this same concept is shown in Figure 10.3. This enables us to concentrate solely on savings and investment and to see the effect of changes more clearly.

Remember that investment is not regarded as directly dependent on the level of income, and so is represented by a line parallel to the national income axis. However savings are dependent directly on income levels, and can be expected to rise as incomes rise: the savings curve is thus shown as positive sloping. Of course this slope must be less than 45°, because such an angle would indicate that each additional £1 of income was entirely saved – an unlikely situation.

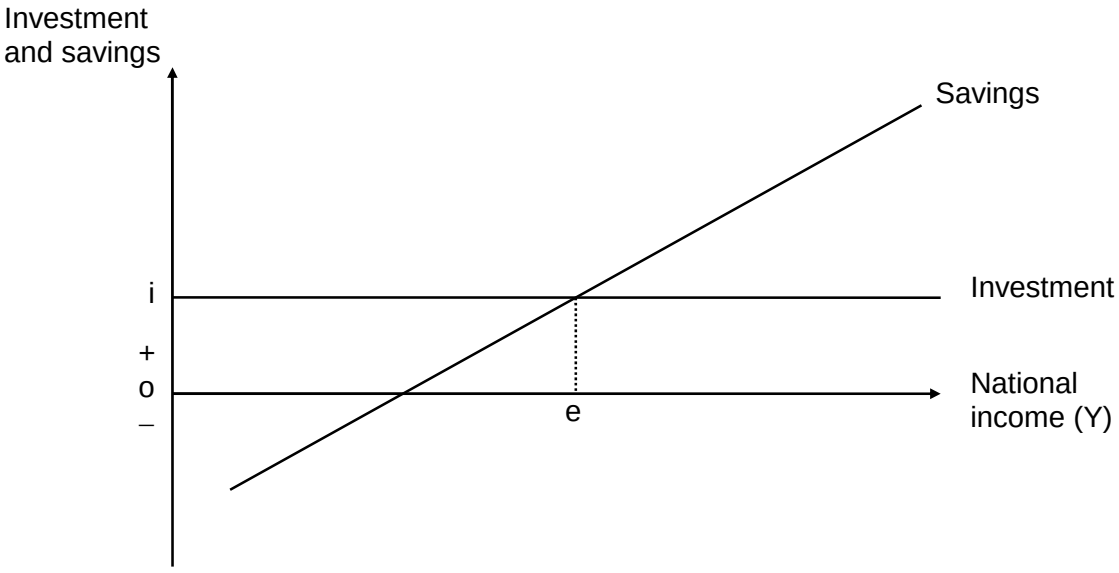


Figure 10.2: National income in equilibrium

Once again, we see that there is one income level where savings will just equal investment, and this is the level that national income will tend to move towards. This is shown as Oe in Figure 10.2. Actual savings will tend to equal actual investment, even though the savings *intentions* of households and the investment *intentions* of business firms are not the same. Remember that it is consumption that tends to bring them together. Firms will seek to

© ABE and RRC

"produce for consumption" that level of output which they believe households will "buy for consumption". Remember too that prices, and stock levels, may change as national income moves into equilibrium.

Now let us see what happens when there is a change in the level of investment. Look at Figure 10.3. Here investment rises, at all income levels, from Oi to Oi_1. As a result, we see that the equilibrium level of income, where actual investment equals actual savings, moves up from Oe to Oe_1.

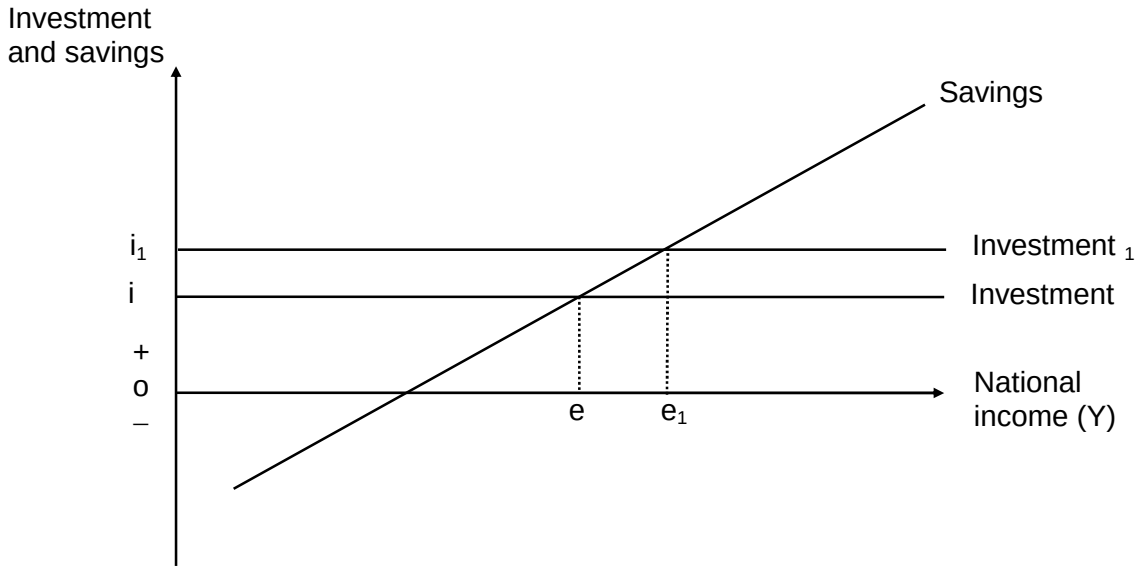


Figure 10.3: A rise in investment

Change in Investment and Change in National Income

We shall now examine the relationship between a change in investment, as just described, and the change in total national income which results from the new equilibrium level. Look now at Figure 10.4.

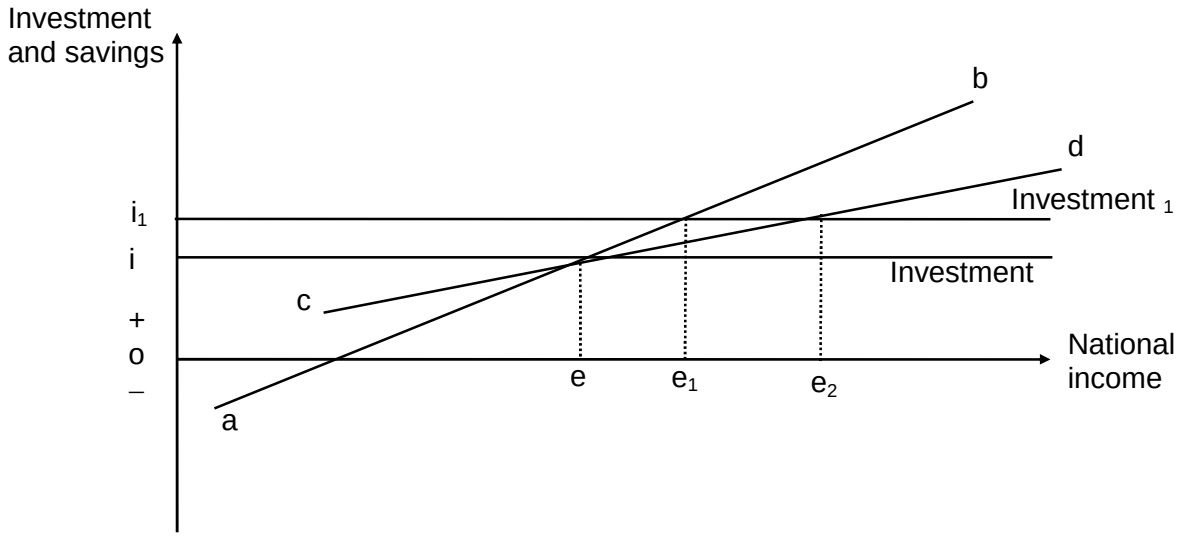


Figure 10.4: Increase in the level of investment at all income levels

This shows an increase in the level of investment at all income levels, from Oi to Oi_1, but now we have two savings curves – ab and cd. Given the savings curve ab, the increase in

investment lifts the equilibrium level of national income from Oe to Oe_1, but, if the savings curve is cd, then the same increase in investment produces the larger income increase from Oe to Oe_2.

We can now state the following.

- An increase/decrease in investment will increase/decrease the equilibrium level of national income.
- The amount of increase/decrease in national income brought about by the change in investment will depend on the slope of the savings curve – i.e. on the amount of any increase in income which is saved.

The more acute the angle of the savings curve, the less is the increase in savings from each additional £1 of income. What is really being represented in this diagram is the multiplying effect of an initial increase in business investment. Suppose that firm A decides to buy an additional machine. This stimulates activity from the machine manufacturer, who increases production and pays additional incomes to his workers. In turn, the workers decide to increase their spending, which stimulates more activity from other firms, and so on. We can visualise successive "rounds" of increased activity, but as some part of each "round" of extra income is saved, the next round is slightly smaller than the last, until the increases become too small to be significant, and the progression comes to an end.

The less the amount saved, the greater will be the total increase. For example, suppose there is an initial increase of 100. The following table shows how this may be multiplied.

In column A, three-quarters of each extra round of income is consumed and one-quarter saved, and in column B, four-fifths is spent on consumption and only one-fifth saved.

	A *(savings 1/4)*	**B** *(savings 1/5)*
Initial	100	100
2nd round	75	80
3rd round	56	64
4th round	42	51
5th round	32	41
Total so far	305	336

Table 10.1: Effect of different rates of saving

These figures are rounded. If we were to produce completely accurate figures and carry on the tables, we would find that A would arrive at a total of 400 and B at a total of 500. Using a calculator, you can test this for yourself. These figures should suggest something to you.

- An initial increase of 100, increased by successive additions of three-quarters, arrives at a final total of 400.
- An initial increase of 100, increased by successive additions of four-fifths, arrives at a final total of 500.

Putting this another way, if the amount saved or held back from each successive increase is one-quarter, then the initial increase is multiplied by four; if the successive increases are reduced by one-fifth, the initial increase is multiplied by five. It looks as though the multiplying effect is the reciprocal of the amount held back from each successive increase. Indeed this is the case.

© ABE and RRC

The Investment Multiplier

This is the term given to the ratio of the change in income to any given change in the level of investment when national income equilibrium has been restored. In symbols, this can be expressed very simply as:

$$K_i = \frac{\Delta Y}{\Delta I}$$

where:

K_i is the investment multiplier

ΔY is the change in national income and

ΔI is the change in investment.

The value of the investment multiplier is the inverse of the amount of each successive increase in income which is saved:

$$K_i = \frac{1}{s} = \frac{1}{1-c}$$

where:

s = proportion of extra income that is saved and

c = proportion of extra income that is spent on consumption.

A more correct definition of s and c would really be the "marginal propensity to save" and the "marginal propensity to consume".

More Realistic Multiplier

So far, we have considered the multiplying effect only in terms of investment and savings, having assumed that the government spends only its taxation revenue and that total exports equal total imports. These assumptions are rather unlikely in modern industrial economies, so a more realistic (and much smaller) multiplier has to take these injections and withdrawals into account.

We can show this in Figure 10.5. This shows an increase in total injections (investment, government spending and exports) and a withdrawals curve. The total withdrawals from income are made up of savings, taxation and imports, so that the propensity to withdraw (w) is now the total of the propensities to save, to tax and to import:

$$w = s + t + m$$

This more realistic multiplier is the ratio of the change in national income to the change in injections which brought it about, and it is the inverse of the propensity to withdraw:

$$K = \frac{\Delta Y}{\Delta J} = \frac{1}{w} = \frac{1}{s+t+m}$$

Suppose that, out of each additional £1 of national income, £0.1 is saved, £0.3 is taxed and £0.2 spent on imports. Then:

$s = 0.1$; $t = 0.3$; $m = 0.2$

so

$$w = (s + t + m) = 0.6$$

K then is 1/w which here is 1/0.6 = 1.67. This is a very much smaller value than the investment multiplier which, in this example, would have been 10.

© ABE and RRC

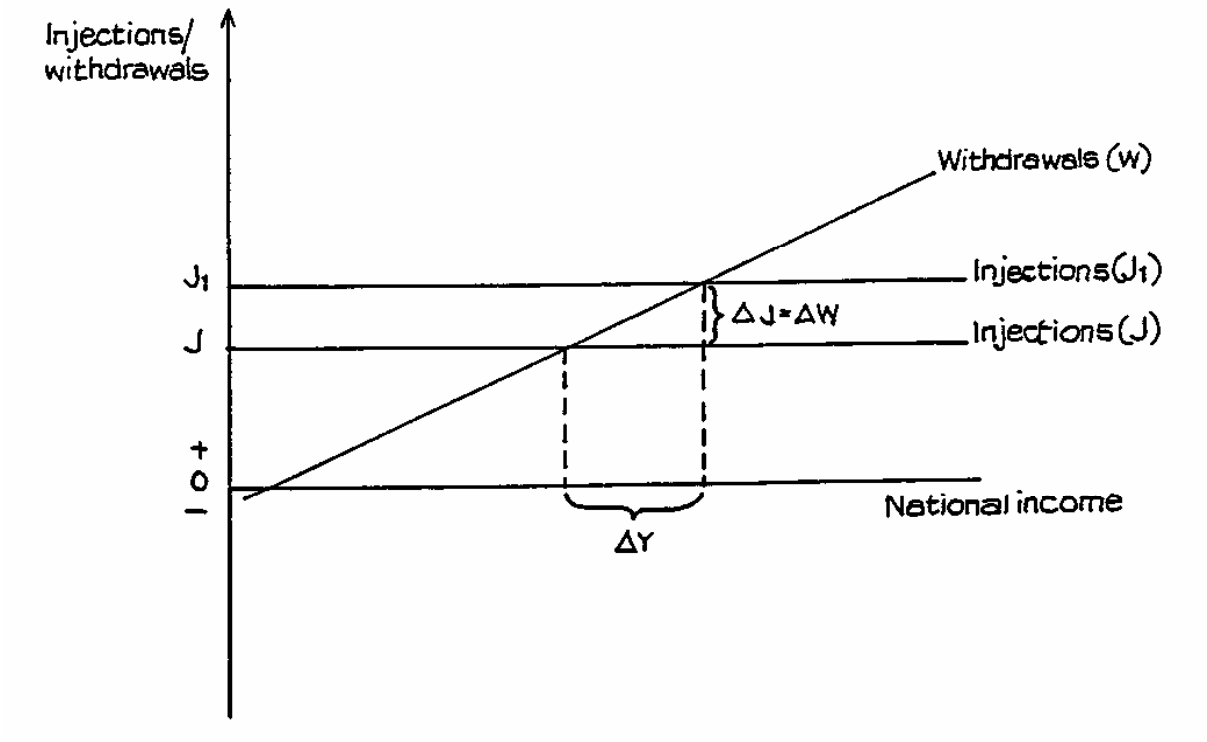


Figure 10.5: Effect of injections and withdrawals

Change in the Marginal Propensity to Save and the Paradox of Thrift

The slope of the savings function (curve) depends on the marginal propensity to save. If people start to save a smaller proportion of their incomes, i.e. spend a higher proportion, then the curve becomes less steep as each additional £1 of income gives rise to a little less saving. If they start to save a larger proportion, i.e. spend a smaller proportion of income, then the curve becomes steeper, subject to a maximum of 45° if the scales on both axes are the same, because each £1 of additional income produces a larger amount of saving though not, we assume, more than the extra income.

This observation has given rise to what has become known as the *paradox of thrift* which is that the more a community tries to save the less it may actually save. This paradox is illustrated in Figure 10.6.

The original equilibrium condition of the national income is represented by Oe_0 where the level of investment and savings are represented by I_0 and Os_0 respectively, i.e. the level where the saving function S_0 intersects the investment level of I_0. Then, for some reason such as a growing fear of unemployment and economic recession or misguided government policy trying to encourage greater "thrift" and "good housekeeping" in the community, people generally start to save more and spend less from their incomes. The saving function becomes steeper and moves, say, to S_1. The equilibrium level of national income falls to Oe_1. Business firms face declining sales and rising stock levels so they cut back their production and invest less in productive equipment. The level of investment falls to I_{t+1}. At this lower level the national income falls further to the equilibrium level where Os_{t+1} equals I_{t+1} at Oe_{t+1}. At this new equilibrium the level of saving has also fallen to Os_{t+1}.

© ABE and RRC

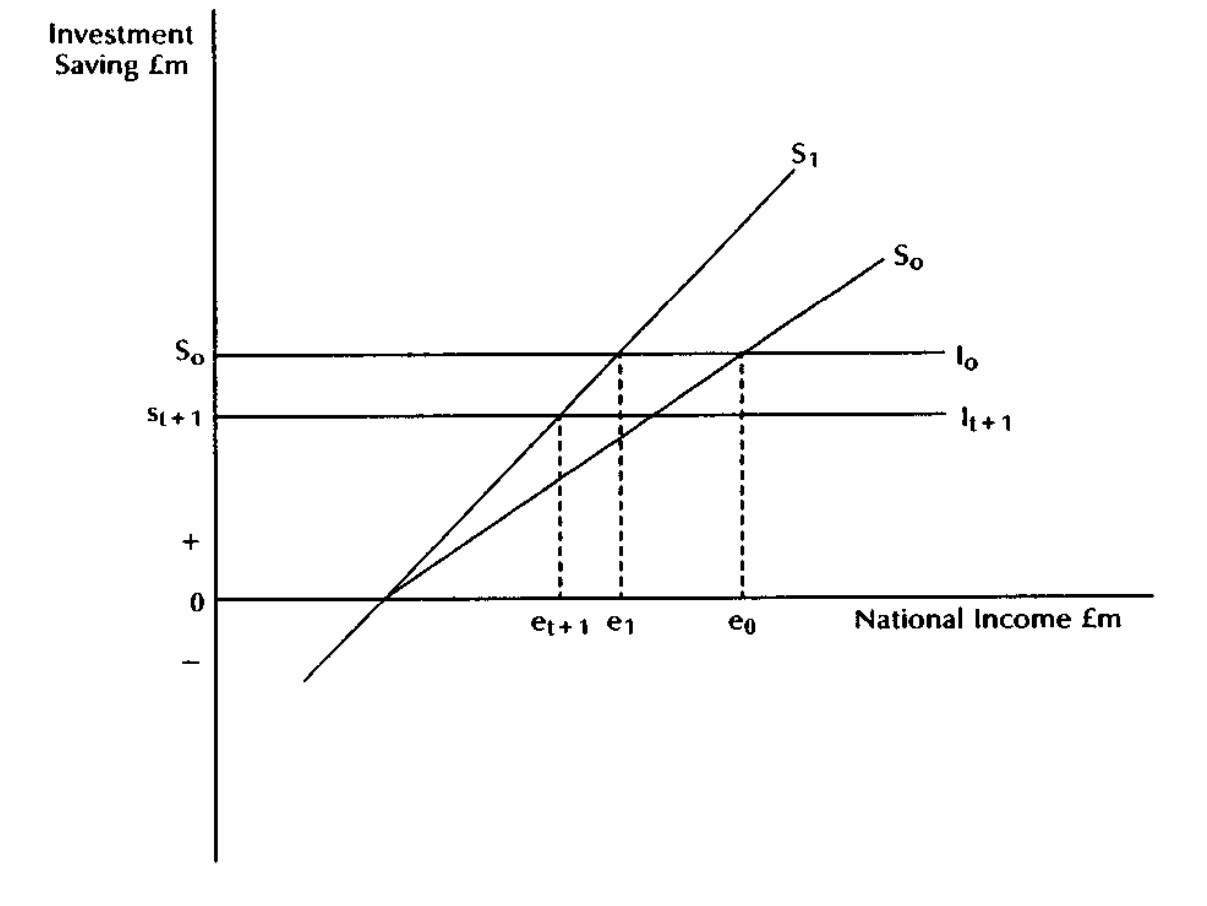


Figure 10.6: Paradox of thrift

Thus, the attempt by the community to save more has resulted in the community actually saving less, because the total level of aggregate income has fallen. Remember this is the result for the community as a whole. Some individual households will have increased their savings, but others will be saving less because they have suffered loss of income and may well be unemployed as a result of the fall in national income and aggregate investment. This is the paradox of thrift in action. This is one case where the macroeconomy (the economy as a whole) behaves differently from the microeconomy (individual firms and households). A virtue for the individual is not necessarily a virtue for the whole community, a concept that some influential politicians have found difficult to grasp.

This example also illustrates the possibility that the fear of recession can become self-fulfilling. If people anticipate that their incomes are likely to fall in the future and start to save more and consume less, their actions can lead to reduced production, investment and employment.

The Investment Accelerator

We have seen that an increase in national income can be induced by a net injection, made up of an increase in the combined forces of investment, government spending and exports. However, if we return to the case of a country in which the government believes in a "balanced budget" (will not spend more than its taxation revenue), where international trade is depressed and there is unlikely to be any net increase from international trade, then we are again left with investment as the main motivating force, other than consumer demand itself.

Now, suppose people do start to consume a higher proportion of their incomes for some reason (the savings curve swings to the right, as in a move from ab to cd in Figure 10.4).

© ABE and RRC

Consumer demand therefore starts to rise. Suppose also that, at the old level of consumer spending, all business equipment was fully used. If business firms believe that consumer demand is on an upward trend, they will wish to increase their productive capacity: to do this, they need to purchase more equipment. There is thus an increase in productive investment.

The principle underlying the theory of the investment accelerator is that there is a constant ratio of investment capital to the total output that is produced, and that this ratio is greater than 1:1. If total demand (and therefore output) is constant, firms will only invest to replace worn-out equipment. However when demand rises, firms will replace old equipment and purchase new, so that the increase in investment is greater than the rise in output desired to meet the rise in demand. But investment will only continue to increase if demand and the output it encourages goes on increasing at a faster rate. If the rise in demand levels off or if demand falls investment will stop increasing or fall. The precise changes to investment will depend on the ratio of investment capital to output and on any time lags between observed changes in demand and business investment decisions.

We have seen that this increase in investment will itself have a multiplying effect on national income, and hence on consumer demand. Initially, the expectations of business firms become self-fulfilling, as their own investment induces the expected rise in consumer spending. Moreover, a quite modest increase in initial consumer spending can have a very great effect on investment spending, as the following rather simplified example will illustrate.

Example:

Let us assume that one machine in the shoemaking industry is capable of producing 10,000 pairs of shoes in a year, that the life of a machine is ten years, and that the industry uses 100 machines, producing a total of 1,000,000 pairs of shoes per annum. Each year, one-tenth of the machines will have to be replaced, so there is a *demand* for ten new machines a year.

What will happen if the demand for shoes increases by 10 per cent? This increase in demand means that 1,100,000 pairs of shoes will be required, and this means that 100 machines must be *used*. The industry will therefore *order* for this year 20 new machines – 10 in order to replace those worn out, and 10 additional ones to cope with the new demand. The demand for machinery will thus *increase* by 100 per cent because of a mere 10 per cent increase in demand for consumer goods.

It is the surge in increased investment spending that gives the accelerator its name.

However there is a danger here. If consumption continues to rise at a constant rate, then investment, after the initial burst, will stay the same. In order that net productive investment should increase, consumption has to continue to increase at a faster rate. If it starts to level off, then investment will fall away. Firms do not need to buy more machines if their production capacity is sufficient to cope with expected demand. A fall in net investment now starts the accelerator in reverse – it becomes a decelerator, forcing a decline in national income. This decline has been caused by nothing more than a levelling of demand and a consequent halt in new business investment.

The Business Cycle

We now have an explanation for the periodic tendency for an economy to expand and decline – to boom and become depressed – which has been a feature of all industrial economies. This cyclical tendency for boom and depression has been described as "the business cycle". Notice that it is explained in terms of consumer demand and business investment, and it assumes that the government is neutral – pursuing a policy of keeping a balanced budget.

The accelerator assumption of a fixed investment capital to output ratio has been criticised on the ground that it very much oversimplifies the business demand for investment, and

© ABE and RRC

ignores a number of important and relevant influences. These include the pace and nature of technological change, competition from foreign producers and changes in the management and use of labour. All these can change the capital to output ratio and the desire to invest at any given time. The basic theory also assumes that firms typically operate at full machine capacity, whereas most of the evidence suggests that it is more normal for firms to operate with some spare capacity, which is used to even out fluctuations in investment. The theory also ignores the influence of the capital market which can have a major effect on the volume and timing of investment. For these and many other reasons earlier hopes that the theory would provide the key to smoothing out the business cycle have proved much too optimistic.

D. THE ROLE OF THE GOVERNMENT IN INCOME DETERMINATION: THE GOVERNMENT'S BUDGET POSITION AND FISCAL POLICY

Although it is helpful to examine the model of income determination and the multiplier process using diagrams, it is also possible to express the model in equation form and solve the equations to determine the equilibrium level of national income. Let us consider the case of a closed economy for simplicity. That is, the economy does not trade with the rest of the world, so that we can ignore imports and exports in the circular flow of income.

We can describe the model of the closed economy with government as follows:

$$Y = C + I + G$$

$$C = 50 + 0.8Yd$$

$$I = 500$$

$$G = 1000$$

Here Y refers to national income, and C, I and G refer, respectively, to consumption, investment and government expenditure. Yd refers to disposable income and the tax rate (t) on income in the economy is 20 per cent or 0.2. That is, instead of simply assuming that government taxation is a fixed sum of money for the whole economy, we have made the much more realistic assumption that the government sets the rate of income tax and its total tax revenue is an increasing function of income. Combining the above equations we can solve for the equilibrium level of national income as follows:

$$Y = 50 + 0.8(Y - 0.2Y) + 500 + 1000$$

$$Y = 0.8Y - 0.16Y + 1550$$

$$Y = 0.64Y + 1550$$

$$Y - 0.64Y = 1550$$

$$0.36Y = 1550$$

$$Ye = 1550/0.36 = 4305.5$$

The formula for the multiplier K is now:

$$K = \frac{1}{1 - c(1 - t)}$$

where c is the marginal propensity to consume and t is the marginal rate of income tax.

© ABE and RRC

Inserting the values given above for the model we obtain:

$$K = \frac{1}{1 - 0.8(1 - 0.2)}$$

$$K = 1/(1 - 0.64)$$

$$K = 1/0.36 = 2.78$$

Now that we have "solved" this system of equations (our model for the equilibrium level of national income and the value of the multiplier) we can see how much more efficient it is to use this approach than relying on diagrams. Suppose the economy is facing a downturn in demand in the economy, due to falling overseas demand for its exports; then the government may decide to boost demand by increasing its own expenditure in the economy, (increasing its injection, G). What would be the consequence of an increase in the level of government expenditure of 500?

The change in income is found as follows:

$$\Delta Y = 500 \times K = 500 \times 2.78 = 1390$$

And the new, higher, equilibrium level of national income is:

$$Ye = 4305.5 + 1390 = 5695.5$$

Such analysis is important to governments seeking to understand the workings of their country's economy and manage the level of aggregate demand for the public good. It is also important for business decisions. For example, suppose a foreign firm is just about to start investing in a new factory to produce consumer goods in an economy, like the one described in the previous simple model: it sees that the country's exports are declining and that the economy is likely to go into recession with rising unemployment. This would clearly not be a good time to invest in the country because the new factory would find it difficult to meet its planned sales targets if the economy was going into recession. The foreign firm might decide to delay (or worse) cancel the building of the factory. However, if it learns that the government is going to increase its spending, and its budget deficit, by an additional 500 to offset the fall in demand due to declining exports, the foreign firm can work out that this will boost demand in the economy by 1390 and lead to an increase in employment. In this case the firm is likely to decide to go ahead with its decision to build the new factory.

In practice of course nothing is ever this simple. If increased government expenditure alone could cure unemployment and increase national income, there would be no poor countries in the world! The limitations of government in the economic management of the economy through fiscal policy are considered in the next study unit.

© ABE and RRC

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved those learning objectives covered in this unit. If you do not think that you understand these objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. What condition is satisfied in the economy when $C + I + G + X = C + S + T + M$?
2. What is the investment multiplier?
3. What is the formula for the simple investment multiplier?
4. All other things remaining unchanged, how will an increase in the marginal propensity to save affect the equilibrium level of national income?
5. You are given the following information about a closed economy:

   $Y = C + I + G$

   $C = 50 + 0.8Yd$

   $I = 500$

   $G = 1000$

   where Y refers to national income, and where C, I and G refer respectively to consumption, investment and government expenditure. Yd refers to disposable income and the tax rate (t) on income in the economy is 20 per cent or 0.2.

   (a) Calculate the equilibrium level of national income.

   (b) Calculate the value of the multiplier for this economy.
6. With reference to the economy described by the equations in question 5, what would be the new equilibrium level of national income if the government increased its level of expenditure from 1000 to 2000?
7. You are asked to give advice to an overseas businessman who is considering investing in the economy described in questions 5 and 6. How will the government's announcement that it is going to increase its expenditure affect the businessman's decision to invest in the economy?

© ABE and RRC

Study Unit 11

Macroeconomic Equilibrium and the Deflationary and Inflationary Gaps

Contents ***Page***

A. National Income Equilibrium and Full Employment 196
Earlier Views – Equilibrium Produces Full Employment 196

B. The Basic Keynesian View 196

C. The Deflationary Gap 197
Possible Causes 197
Consequences 198
Policy Options for Closing the Deflationary Gap 198

D. The Inflationary Gap 200
Some Possible Causes 200
Consequences 202

E. The Aggregate Demand/Aggregate Supply Model of Income Determination 203
Aggregate Demand and Supply 203
The Aggregate Demand Curve 203
Aggregate Supply 204
The Long-Run Aggregate Supply Curve 204
The Short-Run Aggregate Supply Curve 206
The Equilibrium Level of Real Output and the General Price Level 207
Excess and Deficient Aggregate Demand 208
Using Fiscal Policy to Correct a Deficiency of Aggregate Demand 210

F. Financing Fiscal Policy: Budget Deficits and Public Sector Borrowing 211
Financing of the PSBR 212
The General Government Financial Deficit 213
Importance of Public Sector Borrowing 213

G. The Limitations of Fiscal Policy 214

© Licensed to ABE

Objectives

The aim of this unit, in conjunction with Study Unit 10, is to explain the determination of the equilibrium levels of national income using the Keynesian macroeconomic model in a closed and open economy and demonstrate how this can be of use to businesses.

When you have completed this study unit *and Study Unit 10* you will be able to:

- interpret, graph, and solve simple numerical examples of the form $Y = C + I + G + (X - M)$
- explain how variations in the marginal propensity to save, consume, and import affects the closed and open economy multiplier
- compare and contrast inflationary and deflationary gaps using Keynesian cross diagrams
- discuss the components of government fiscal policy and explain how changes in these components affect the equilibrium level of national income
- make judgements about the factors that determine the effectiveness of fiscal policy
- explain the implications of fiscal policy for government borrowing (Public Sector Borrowing Requirement).

A. NATIONAL INCOME EQUILIBRIUM AND FULL EMPLOYMENT

Earlier Views – Equilibrium Produces Full Employment

Earlier classical economists appreciated the concept of national income equilibrium but believed that, if the economic forces were left to work freely, this equilibrium level would also produce a situation of full employment. They argued that as incomes fell, labour costs would also fall, until it became worthwhile for business entrepreneurs to increase their demand for workers.

If instead of this happening there was large-scale unemployment, then it was argued that the fault lay with trade unions and other institutional forces in the economy: they were keeping up wages and prices and making labour overpriced in relation to the current level of demand. The remedy for unemployment lay in forcing down wages despite any opposition that might be encountered.

B. THE BASIC KEYNESIAN VIEW

Keynes accepted that, in the long run, it might be possible to bring down wages until labour became so cheap that all workers wanting jobs could be found employment. However, he regarded the price of such action, in terms of social distress and political conflict, as being unacceptable in a modern society. He doubted whether society could withstand the conflicts and pressures that would be set up by the attempt to bring down wages far enough to achieve full employment.

Therefore for practical purposes, and in the interests of social and political peace, he considered that it was better to regard the equilibrium level of national income and the level at which all workers were fully employed as two separate levels, with no natural way of coming together through the operation of the normal economic forces.

This concept of the separation of equilibrium and full employment levels of national income is illustrated in Figure 11.1. Here, we return to the model based on the 45° line which, you

© Licensed to ABE

will remember, represents the curve where all income is expended. The intended levels of expenditure at each level of income are shown by the curve C + J (consumer spending plus total injections from investment, government and exports).

The equilibrium level, where intentions are fulfilled without changes in prices and stocks, is Oe, where the C + J curve intersects the 45° line.

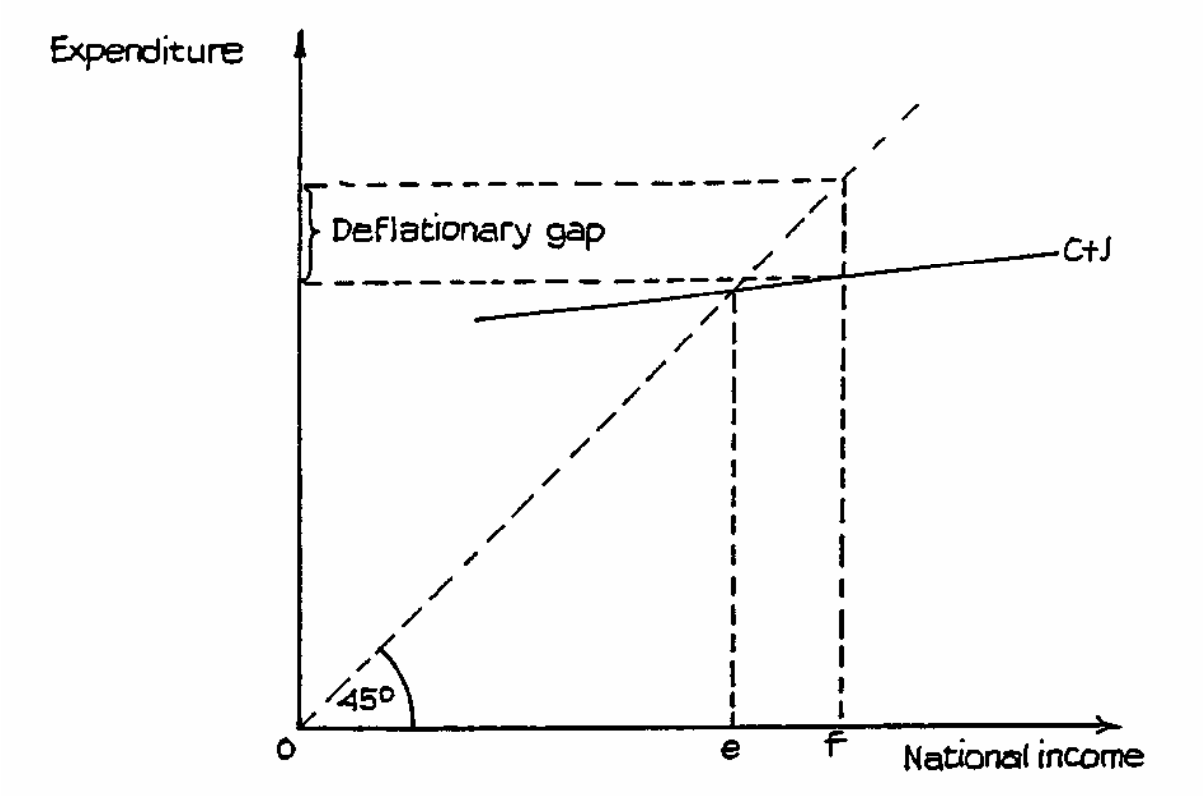


Figure 11.1: The separation of equilibrium and full employment levels

Suppose that possible output of goods and services available for purchase by the community, given full employment of all those seeking work, would push up income to level Of. However, at this level of income there is a gap between the 45° line and the C + J curve. This gap indicates that possible expenditure at this income level is greater than intended spending from the total forces of consumption, investment, government and exports.

C. THE DEFLATIONARY GAP

The basic model of the deflationary gap was shown in Figure 11.1. The gap arises when total aggregate demand from household consumption, business investment, government spending and net exports (C + I + G + (X − M)) is insufficient to absorb all the output that could be produced if all available production factors, including those workers seeking employment, were fully employed.

Possible Causes

Strict classical and monetarist economists believe that a deflationary gap would not exist if both product and factor markets were free to perform their basic functions of bringing supply into equilibrium with demand through changes in price. Closing a gap by the operation of market forces alone would imply significant reductions in wages. However wage incomes are a major influence on the level of consumer demand, so that any large-scale reduction in wage levels would further depress the consumption element in aggregate demand. Fear of

© Licensed to ABE

future unemployment and falling incomes would also depress demand and of course business investment, so that there is no guarantee that greater wage flexibility in a modern economy would close the gap. It could make it larger. Actions of business firms in making workers redundant, and deliberately creating an atmosphere of insecurity in their workforces to keep wage levels restrained, could be one of the initial causes of the deflationary gap.

Government action to reduce spending and to reduce the size of the public sector in the economy could have a similar effect, both in reducing the G element in aggregate demand and in undermining consumer and business confidence in the future of the economy, and so causing the gap and then making it wider.

Consequences

The immediate and most visible consequence is a rise in unemployment and lengthening of the time that the unemployed remain out of work. This is the feature that made the Great Depression of the 1930s such a searing experience for all those who experienced it. It shaped economic and political attitudes for a generation, until memories of the depression became submerged beneath the more recent and longer-lasting experience of inflation. Long-term unemployment creates severe social and personal problems, as well as being a cruel waste of potentially productive economic resources. In Keynesian thinking it is something that governments can and should seek to remedy and preferably avoid altogether.

However, labour is not the only factor of production. In a severe economic depression all factors are unemployed or underemployed. Land goes out of cultivation, business premises remain empty and deteriorate, and machines lie idle and rust. If supply is greater than demand in the factor and major product markets we would expect prices to fall. In some markets, notably the private house and business property markets, there have been price reductions. However, property is regarded as a form of wealth rather than as a consumer good, and price reductions for private houses are not welcomed by households in the way that price reductions for, say, furniture or private cars would be welcomed. People feel poorer when the value of their home falls, especially if they have a mortgage loan that is larger than the home's current market value (negative equity). Under these conditions home movements and associated purchases are much reduced and in general consumer spending is depressed.

Policy Options for Closing the Deflationary Gap

The implication of the basic Keynesian model of the deflationary gap is that the aggregate demand curve of $C + I + G + (X - M)$ or $C + J$ (J standing for all the demand injections) should be raised to bring the equilibrium level of national income closer to the full potential employment level. This is illustrated in Figure 11.2.

© Licensed to ABE

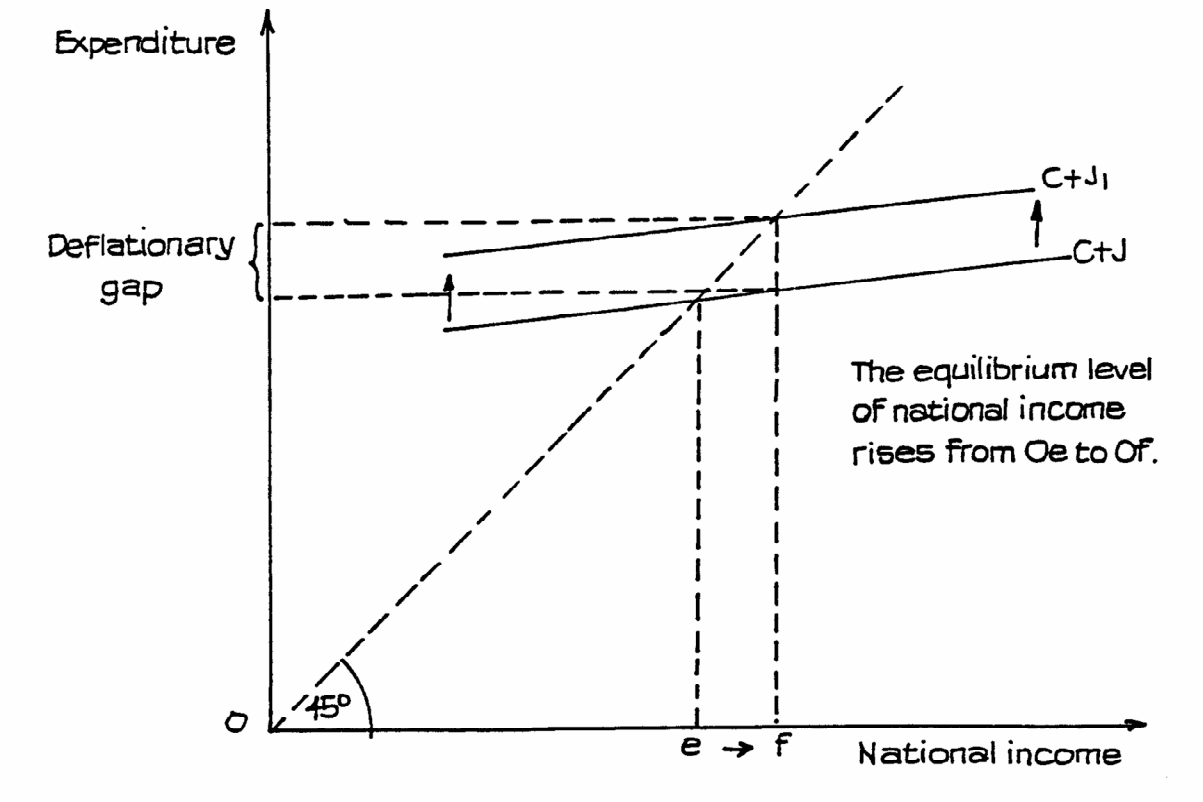


Figure 11.2: Raising the aggregate demand curve

Since business investment (I) levels are a consequence of firms' experience of past and current consumer demand, and their view of the probable future trend of this demand is also dependent on net export levels, the potential for lifting I when C is depressed is limited. However, there is one other element within total aggregate demand which is not necessarily an inevitable part of the business cycle: government spending (G). Government spending is the result of political decisions that can be taken independently of the national income and consumer demand, *if* the government abandons the principle of the balanced budget (spending equals taxation revenue). This of course is government spending on such projects as road and communications development.

The possible result of increasing government spending is shown by the movement in the C + J curve in Figure 11.2. Here, we see that the rise from C + J to C + J_1, brought about by an increase in government spending, is able to close the deflationary gap and remove large-scale unemployment. This, very broadly, was the type of remedy advocated by Keynes for the massive unemployment problem of the 1930s.

Unemployment in Britain did start to fall when government spending began to increase in the face of approaching war, in the late 1930s. However, a remedy that was developed in the 1930s does not necessarily apply quite so simply in the very different economic conditions of today, and we need to examine the whole problem much more carefully (which we shall do in subsequent study units).

Modern Keynesians now recognise that continued demand stimulation policies, aimed at closing the deflationary gap by accepting an unbalanced budget and relatively high levels of government borrowing, can have inflationary effects leading eventually to the problem of stagflation, when both unemployment and prices rise together.

Keynesians now accept that the demand-management policies of the 1950s and 1960s contributed to the high inflation suffered in the 1970s. Most are prepared to agree that they

© Licensed to ABE

had understated a number of consequences of government measures to keep unemployment low. These included:

- The rapid expansion of the public sector fed by injections of government spending, and the relative contraction of the private sector as this became uncompetitive in world markets. Expansion of public spending beyond the capacity of tax revenues to sustain it led to large amounts of government borrowing. These combined to increase inflationary pressures in the economy.
- Long periods of low unemployment, and a belief that governments would always act to avoid high unemployment, gave labour unions an inflated view of their own power. Union pressure to raise wages and achieve generous legislation to provide job security, in spite of increased competition in world markets, aggravated the problem of stagflation. It delayed the improvement in labour productivity that was needed to increase domestic production, and slow down the decline in exports and rises in imports experienced during the 1970s.

Modern Keynesians also recognise that the technological revolution of microelectronics has fundamentally changed the structure of industry, and shifted the long-run labour to capital ratio in modern production in favour of capital. They accept that industrial practices have to become more efficient if firms are to compete successfully in world markets.

At the same time, Keynesians have retained their basic belief in the duty and ability of government to intervene to mitigate the social effects of economic cycles and the consequences of technological change. They do not believe that unregulated markets will always lead to equilibrium conditions acceptable to modern society, and they continue to place importance on the public sector provision of those goods and services that are inadequately provided by private sector markets.

D. THE INFLATIONARY GAP

An inflationary gap is created when aggregate demand of C + J is greater than the supply of goods and services provided when national income is operating at or near the full employment level. Such a gap is illustrated in Figure 11.3.

Here total demand, from all the forces represented by the C + J curve, is forcing an equilibrium level of national income above the level of total production and real spending that is possible given full employment at income level Of. The pressure to buy goods and services that are not being produced forces up prices. In this situation, total spending intentions cannot be fulfilled, so that actual spending is lower than intended.

Some Possible Causes

Keynesian models are better at coping with unemployment than with inflation, and Keynesian economics went into retreat in the face of the massive inflation of the 1970s and 1980s. Earlier Keynesians had been prepared to tolerate a low rate of inflation, perhaps around three per cent, in the belief this provided a stimulus to demand and helped to keep unemployment levels low.

Experience has shown that low inflation rates can very rapidly turn into high rates. The inflationary gap produces price rises and waiting lists for goods and services. Unfortunately these do not actually close the gap. If prices rise, people spend the money they had planned to spend, but do not buy all the goods and services they had planned to acquire. The spending pressure remains high and rising prices actually increase demand, since people prefer to buy now at today's price rather than tomorrow at a higher price. If they finance this spending by borrowing they increase the money supply and this adds further inflationary pressures.

© Licensed to ABE

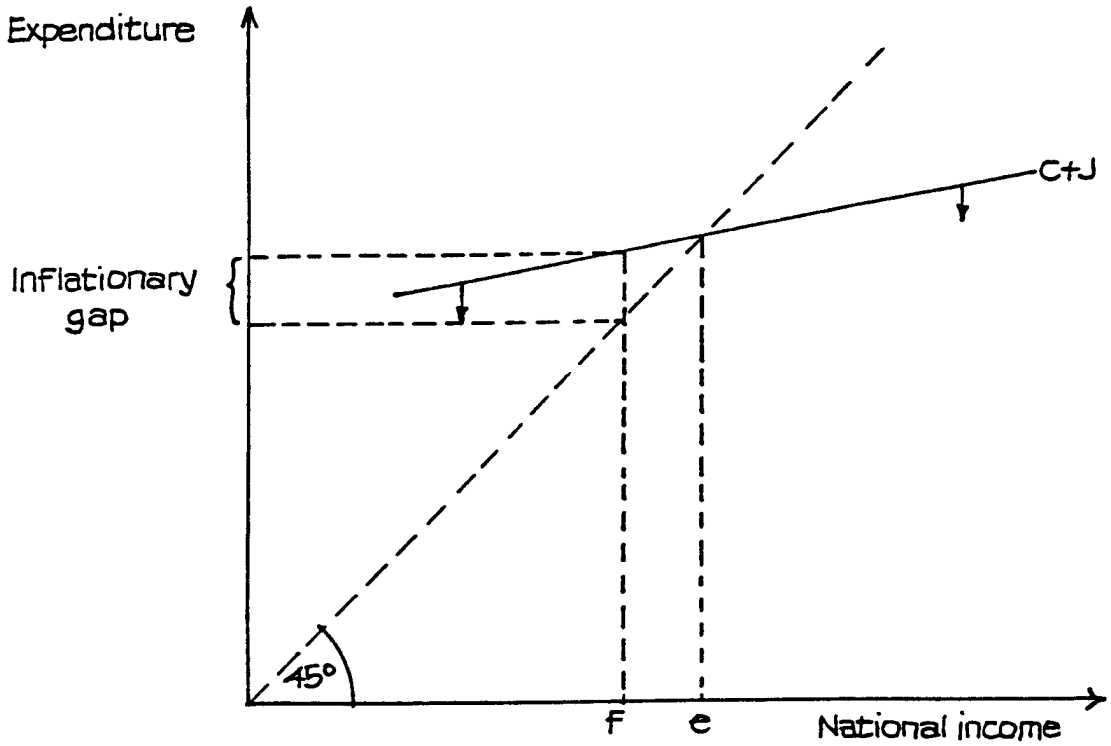


Figure 11.3: The inflationary gap

In its simplest terms an inflationary gap arises when aggregate demand is greater than aggregate supply, which is unable to respond sufficiently to reduce the excess demand. This then raises two questions:

(a) What causes the excess demand?

(b) Why, if it is the function of a market economy for supply to respond to demand, is the production system unable to meet total demand?

In their extreme forms, Keynesians and monetarists have given conflicting answers to these questions. Today, they are closer together, but still place different emphasis on different aspects. At this stage these differences are just outlined.

Keynesians have blamed excess demand on excess income which is running ahead of potential production. More recently, they have been prepared to accept that money supply and government borrowing have also played a significant part in stimulating demand.

Monetarists have tended to blame excessive demand on excess money supply (for reasons that are explained later), but they have also linked this with rising wage levels made possible by business borrowing. They have also linked excess money supply to government spending and borrowing.

The original Keynesian model of the inflationary gap assumed that the production system could respond to rising demand, up to the point where all production factors were fully employed. A significant inflationary gap would only appear when the equilibrium level of national product rose above the full employment level. This basic model offered little scope for a convincing explanation for the stagflation of the 1970s and early 1980s, when both inflation and unemployment were rising. Consequently Keynesians have had to accept deficiencies in the production system at levels below full employment. As already explained, they have tended to emphasise problems arising from a period of rapid structural change caused by the contemporary technological revolution.

© Licensed to ABE

Monetarists have traditionally been more prepared to see inflation and unemployment as associated, rather than opposing problems of a troubled economy. They do not only regard inflation as a cause of unemployment because of its effect on business productivity and ability to compete in world markets. They also see inflation as being partly caused by defects in the supply side of the economy that encourage people to remain unemployed even though there is excess demand in the economy. Inefficient factor markets permit unused production capacity to remain unused in spite of high levels of demand. However, they have had to recognise the deflationary and unemployment consequences of their monetary and market reform/supply side policies aimed at reducing inflation. Inflation control has proved a far more difficult economic and social problem than the monetarists anticipated.

Consequences

In the 1950s and early 1960s inflation rates were low by later standards. When the economy was growing at unexpectedly encouraging levels, it was not uncommon for observers to comment that a low rate of inflation might be healthy and stimulating for an economy. However, as explained earlier, inflation tends to feed on itself, and can suddenly rise out of control unless measures are taken to impose checks. The common socio-economic problems arising from inflation have tended to be identified as:

- Countries with inflation rates higher than their trading partners and/or rivals soon price themselves out of increasingly competitive world markets. Exports fall and imports rise, so that an international payments problem undermines the currency (in ways that are discussed more fully in a later unit). To this extent inflation leads to rising unemployment.
- Confidence is lost in the stability of the domestic currency and financial structure. Savings fall and there is a flight of capital in spite of any financial exchange controls that might be imposed. In extreme cases a flight from money to physical goods fuels further inflation.
- As long as most incomes rise faster than prices people can be misled by an impression of rising wealth, particularly when high-value fixed assets such as houses and land gain high monetary values. However, as more and more sections of the population fail to maintain the real (inflation adjusted) value of their incomes, and living standards fall for a growing number of people, there is a big increase in social discontent. In extreme cases there is civil conflict, destruction of property and loss of lives. At this stage there is a danger of complete social and political breakdown with unpredictable consequences.

During the period of high and rising inflation of the 1970s there were still those who defended inflation as being preferable to high unemployment. They argued that there would be no undesirable consequences if all financial payments and obligations were to be "index linked", i.e. if all monetary values were periodically adjusted by an agreed inflation measure. Some even argued that this would itself gradually bring down the rate of inflation, since there would be nothing to gain from raising prices when costs also rose at the same rate.

In practice, experience soon showed that although some degree of indexation was able to preserve the value of some obligations, such as real yields on savings and the purchasing power of pensions, inflation itself is too complex and uneven in its effects for it to be simply indexed away into insignificance. It also became clear that the low-inflation countries, such as Germany and Japan, were able to enjoy more successful economic growth and higher living standards than the high-inflation countries such as Britain and Italy. Indexation was, of course, no cure for the international trade problems of the high-inflation countries.

By the 1980s there was widespread agreement throughout Western Europe that inflation was a major economic problem that governments had to solve. There was sufficient popular

© Licensed to ABE

support for this for governments to risk taking measures that they knew would increase unemployment in the short-term.

Indeed, it is now recognised that the Keynesian injections and withdrawals model, represented by the 45° model of the economy, is incomplete. It leads to an over-optimistic picture of the power of fiscal policy to alter permanently the equilibrium level of output in an economy. To understand the nature of this limitation and work with a more realistic model of income determination, we need to relate the level of demand in the economy to the economy's supply capability.

E. THE AGGREGATE DEMAND/AGGREGATE SUPPLY MODEL OF INCOME DETERMINATION

Aggregate Demand and Supply

The major limitations of the Keynesian injections/withdrawals model are that it focuses exclusively on the demand side of the economy, and neglects completely the supply side of the economy. Although the model can be used to illustrate the concepts of an inflationary or deflationary gap, by relating the level of aggregate demand in the economy to its full capacity output level, there is no consideration of the relationship between supply and the price level. The Keynesian model is deficient when it comes to studying the relationship between changes in aggregate demand and the general level of prices in an economy. To understand the causes of inflation and deflation in an economy, and how changes in the level of aggregate demand affect the price level as well as output and employment, a different model is needed. This model is known as the aggregate demand (AD) and aggregate supply (AS) model of income determination.

The Aggregate Demand Curve

An aggregate demand curve is illustrated in Figure 11.4. The aggregate demand curve looks to be the same as the microeconomic demand curve used in earlier units, but appearances can be deceptive. In the aggregate demand and supply diagram the vertical axis in the diagram shows the level of prices in the economy as a whole, and not the price of a single good or service. The price level is measured by an index number of prices, an average measure of all the prices in the economy. This is not the same as the rate of inflation or deflation, but a change in the general level of prices in an economy corresponds to the rate of inflation or deflation. For example, the rise in the price level from P2 to P1 shown in Figure 11.4 implies a positive rate of inflation in the economy. The horizontal axis measures the level of real output or real national income in the economy, not the money value of income or output. Real national income is the measure of output that matters for an economy because it is this that determines the standard of living and the level of employment. If the price of all the goods and services in the economy were to increase by 20 per cent, due to inflation, the value of national output measured in monetary terms would also increase by 20 per cent; but no one would be any better off, because real output would be the same. Actually, if the level of prices rose in an economy due to inflation while all the other economic variables remained the same, the economy would be worse off in the sense that the level of aggregate demand would be lower. This relationship is shown by the downward slope of the aggregate demand (AD) curve from left to right. The downward sloping AD curve results from the fact that as the general level of prices is reduced the real value of the supply of money increases, and the level of the rate of interest decreases. Without explaining this relationship in more detail at this stage, we can deduce that as the general level of prices and the rate of interest decrease, consumers increase their consumption expenditure and firms increase their borrowing and investment expenditure. Thus, all other things remaining constant, as the general level of prices in the economy falls

© Licensed to ABE

the C and I components of aggregate demand increase: the AD curve slopes downwards from left to right as drawn in Figure 11.4.

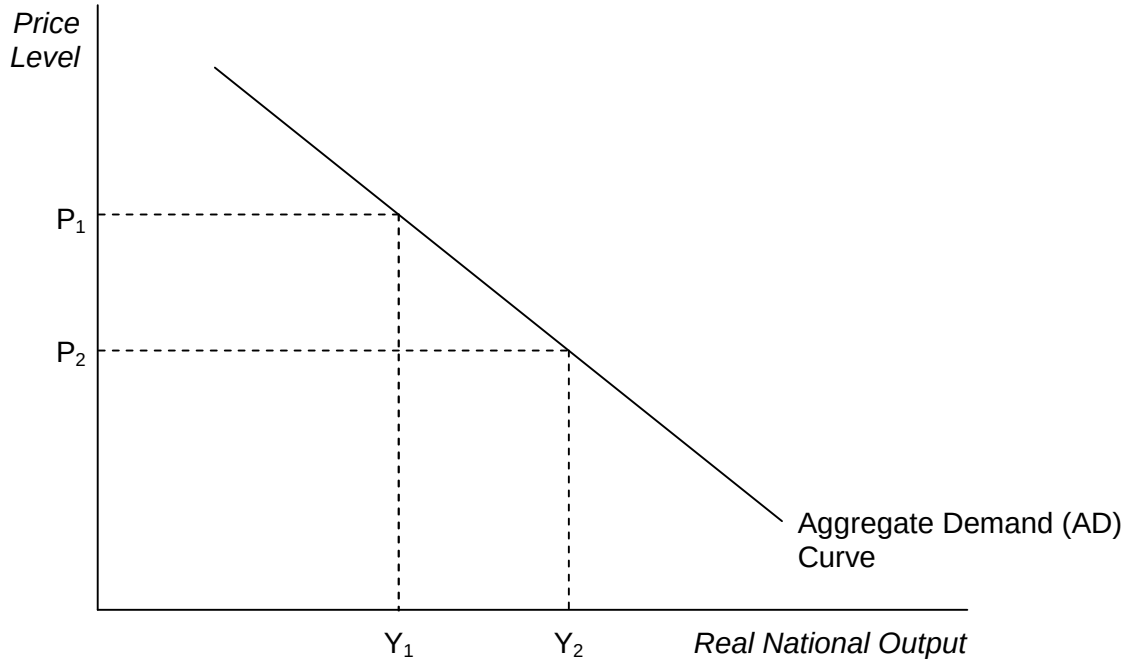


Figure 11.4: Aggregate demand curve

The entire aggregate demand curve will shift to either the left or the right if, without any change in the level of prices in the economy, there is a change in one of the underlying components of aggregate demand or the supply of money in the economy. For example, all other things remaining constant including government tax revenue, an increase in the level of government expenditure will shift the entire AD curve to the right. Conversely, all other things remaining constant, a decision by consumers to spend less on consumption, which is the same as a decision to save a larger fraction of their incomes, will result in a shift to the left in the AD curve.

Aggregate Supply

Aggregate supply (AS) is the economy's total output of goods and services over a given period of time. At the level of the whole economy, we have to recognise that there are two distinct aggregate supply relationships. One is the economy's maximum sustainable level of output. This is termed *long-run aggregate supply* (LRAS). The other aggregate supply relationship shows how the economy can vary its output in the short term, and recognises that for short periods of time it is possible to produce more real output than is sustainable over longer periods. Think of it this way: it is possible for a person to increase their output by cutting down on time spent sleeping and working longer hours each day. However after a few days with little or no sleep, production would fall to zero because of the need to catch up on lost sleep! This kind of relationship is represented by an economy's *short-run aggregate supply* (SRAS) curve.

The Long-Run Aggregate Supply Curve

The LRAS curve is illustrated in Figure 11.5. An economy's level of real national output, and hence the standard of living in the economy, depends upon its natural resources, and its stock of physical and human capital. Provided an economy has the capability to utilise its natural resources effectively, then the greater its endowment of natural resources the higher

© Licensed to ABE

its level of real income. The more and the better the quality of an economy's physical capital, the higher will be its level of real output. An economy's physical capital includes its infrastructure of roads, ports, railways, airports, schools, universities and hospitals, plus all its houses, offices and factories, plus all the vehicles, machinery and equipment. Likewise, the higher the quality of the labour force in terms of education, training and skills, as well as health and life expectancy, the greater will be their productivity and the level of real output in the economy. It is these differences in natural resources, and physical and human capital, that explain the differences in national income and living standards between countries. While the importance of natural resources and physical capital is self evident in explaining differences in income levels between countries, it is differences in human capital that account for the greatest difference in many cases. The efficiency or productivity of the labour force is a major determinant of real national output. This explains why education and training are so important in determining living standards, and why they are given so much emphasis in developed, high income, countries.

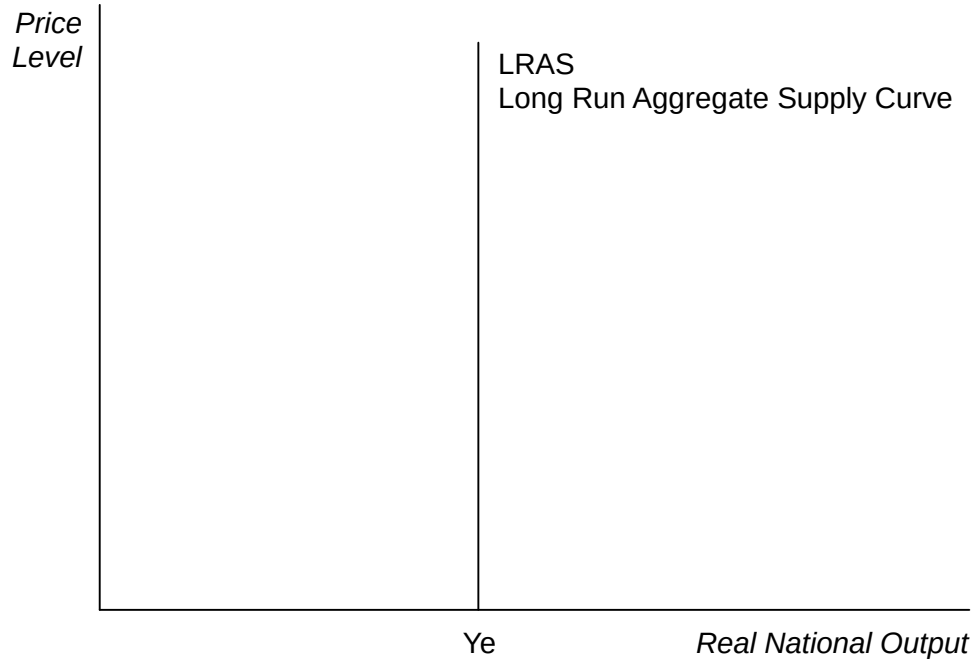


Figure 11.5: LRAS curve

The LRAS curve is vertical at the level of real output determined by the full utilisation of all the economy's factors of production. This point is also termed the point of full capacity utilisation, the point of full employment, or full employment output. The LRAS curve is shown as vertical, that is, completely price inelastic with respect to the general level of prices. This is because once the economy is operating at its sustainable level of full capacity utilisation merely increasing the level of prices in the economy will not result in any increase in real output. Inflation alone does not have the power to make the economy more productive and increase the availability of goods and services.

The position of the LRAS curve is not fixed permanently. Economic growth resulting from increases in the productivity of the economy's factors of production, and/or increases in the available supply of labour and capital through investment, increases the full capacity level of real output. That is, economic growth shifts the LRAS curve to the right. Equivalently, the rightward shift of the LRAS curve through economic growth is equivalent to the rightward expansion of an economy's production possibility frontier. The LRAS curve can also shift inwards to the origin, although fortunately this is much less common, if an economy's

© Licensed to ABE

productive capacity is destroyed through war or natural disaster (such as an earthquake or flooding).

The Short-Run Aggregate Supply Curve

Although an economy cannot maintain a level of total output permanently above that corresponding to its full capacity utilisation output, unless it experiences real economic growth, it can produce to the right of its LRAS curve in the short run. The explanation is simple. Physical capital can be worked for longer periods without maintenance and repair, even if this means that it will breakdown and wear out sooner than its designers intended. Likewise, over short periods of time, workers and land can be worked more intensively and for longer hours than is good for their longer-term health and productivity. However, working an economy's fixed available supply of land and physical productive capital more intensively, by employing more workers and increasing the hours worked, is subject to the law of diminishing returns. This means that for a given level of money wages in an economy the cost of each unit of additional output will rise. Thus the SRAS curve will slope upwards from the left to the right and appear to look just the same as the individual firm and industry supply curves considered in earlier units. An economy's SRAS curve is shown below in Figure 11.6. That the economy's aggregate supply curve, at least in the short run, slopes upward in the same way as a firm's supply curve should not be surprising, because the aggregate supply curve is simply the sum of all the supply curves of individual firms.

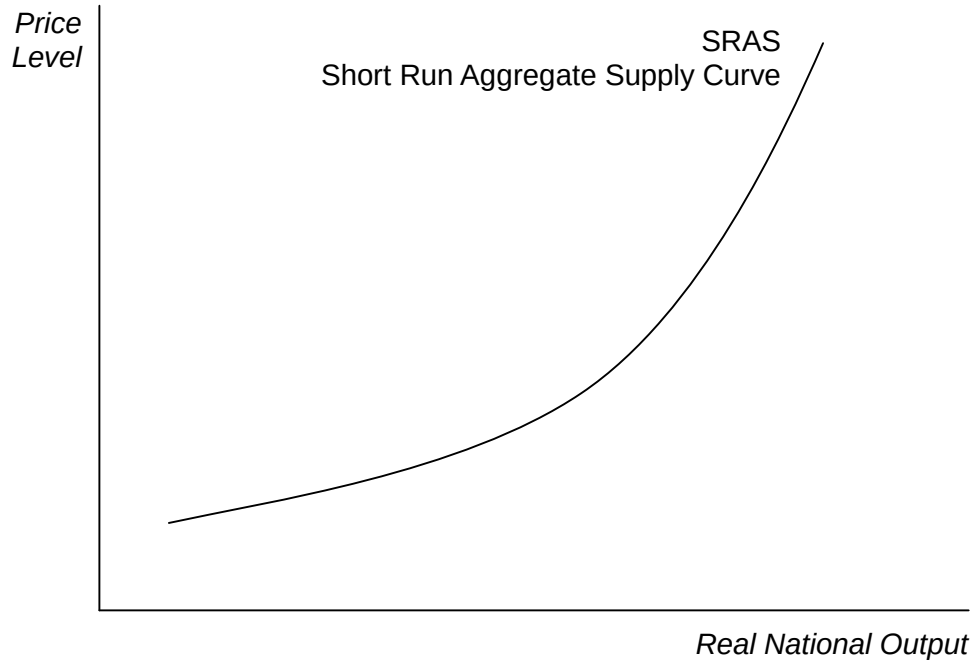


Figure 11.6: SRAS curve

The upward slope of the curve shows that unit costs of production, and hence prices, rise because of diminishing returns as the economy increases its level of real output from a given stock of resources. Each SRAS curve is based upon the assumption of a given level of money wage rates and rates of tax in the economy. Thus, in contrast with the economy's LRAS curve which is fixed at each point in time, there are many possible SRAS curves at any one time depending upon the level of money wages, taxes and import prices. Three such SRAS curves are shown in Figure 11.7.

© Licensed to ABE

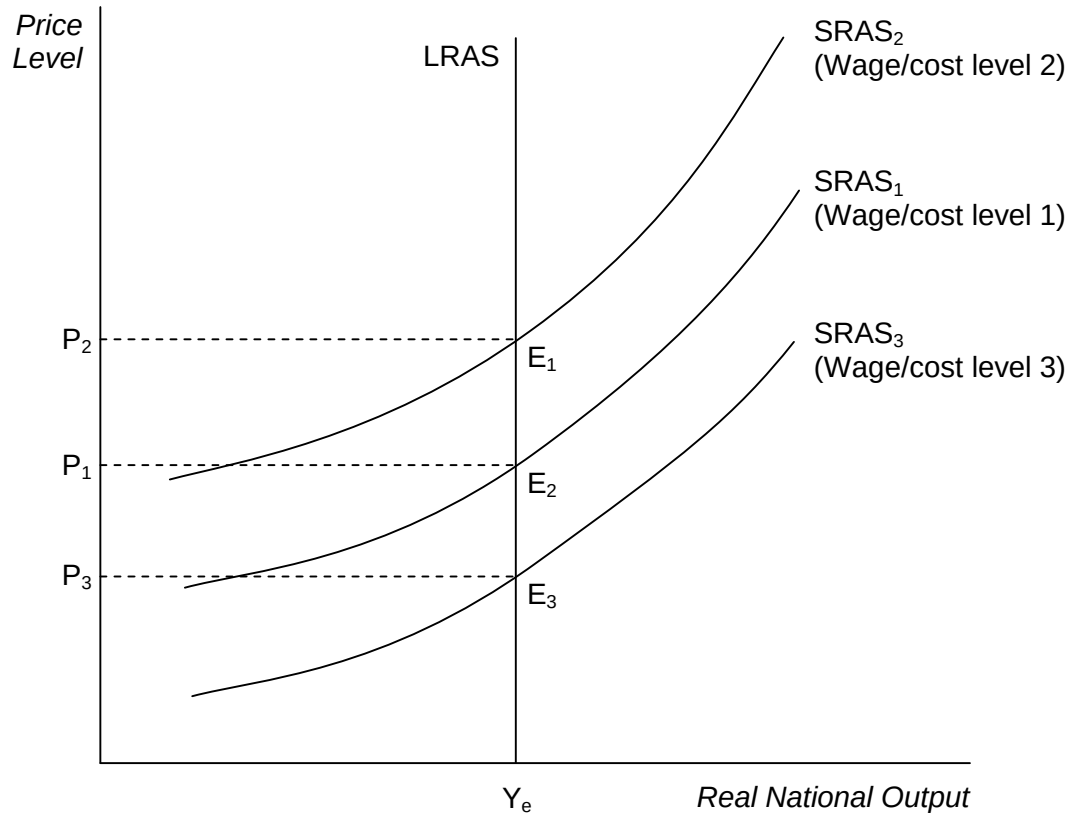


Figure 11.7: A set of SRAS curves

The curve SRAS1 is based upon a given level of money wages. The point of full employment equilibrium is at E_1 where the SRAS curve intersects the economy's LRAS curve. At this point the level of prices in the economy is P_1. Now suppose that there is an increase in the level of money wages in the economy, without any corresponding increase in productivity. This will cause the SRAS curve to shift upwards as shown by $SRAS_2$ in Figure 11.7. At the new point of full employment equilibrium on the LRAS curve, E_2, the level of prices in the economy has increased in proportion to the increase in money wage rates, P_2. This illustrates the fundamental point that simply increasing money wages and other costs in an economy, without any corresponding increase in productivity, will at full employment merely lead to higher prices. The same applies if the increase in costs is due to a rise in the cost of imported energy, such as oil. On the other hand, a reduction in the level of money wage rates in the economy, or a fall in the price of imported raw materials and energy, or a reduction in the level of indirect taxes, will shift the SRAS curve downwards to the right. This is shown in Figure 11.7 by the movement from $SRAS_1$ to $SRAS_3$, and the fall in the general price level from P_1 to P_3.

The Equilibrium Level of Real Output and the General Price Level

The equilibrium level of real national output and the general level of prices in the economy is determined by the interaction of aggregate demand and aggregate supply. The intersection of the AD and SRAS curves determines the economy's equilibrium position in the short run. In the short run the economy can suffer from deficient demand, and be in equilibrium with unemployment, or experience excess demand, over full employment and inflation. If the economy achieves full employment without excess aggregate demand the equilibrium point will lie on its LRAS curve and all three curves must intersect at the same point. This is shown at point E in Figure 11.8.

© Licensed to ABE

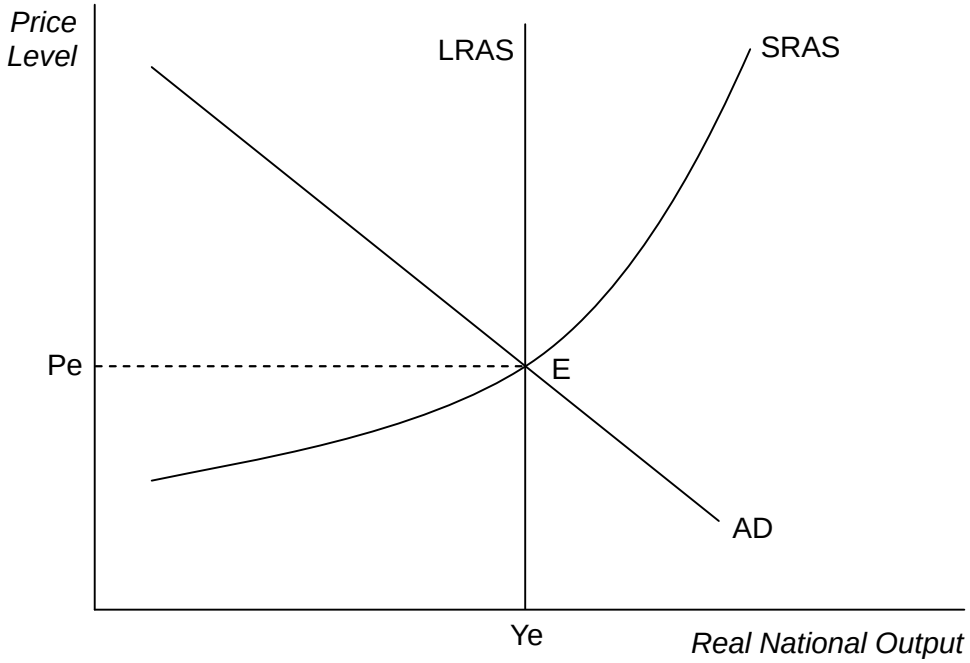


Figure 11.8: Equilibrium level of real output and general price level

Excess and Deficient Aggregate Demand

We have now brought together all the pieces of the aggregate demand and supply model for the determination of the equilibrium levels of real national output and prices. We can use this model to revisit the concept of inflationary and deflationary gaps examined using the Keynesian 45° model earlier in this unit.

In Figure 11.9 the aggregate demand curve AD_1 intersects the SRAS curve at point E_1 to the right of the LRAS curve. This illustrates a situation of excess aggregate demand in the economy and corresponds to the inflationary gap of the earlier analysis. But in the AD/AS model we can see that the point of equilibrium at E_1 is unsustainable because the associated level of real national output, Y_1, is greater than the economy's long-run output level, Ye. The excess demand will place upwards pressure on wages and hence prices in the economy. The SRAS curve will shift upwards with the rise in the level of money wages until a new point of equilibrium is reached at point E_2 on the LRAS curve. The economy will experience inflation as it moves to its sustainable equilibrium at point E_2 with a higher general level of prices in the economy, P_2. Inflationary gaps are essentially self correcting unless the economy experiences a further injection of aggregate demand during the movement to the new equilibrium.

© Licensed to ABE

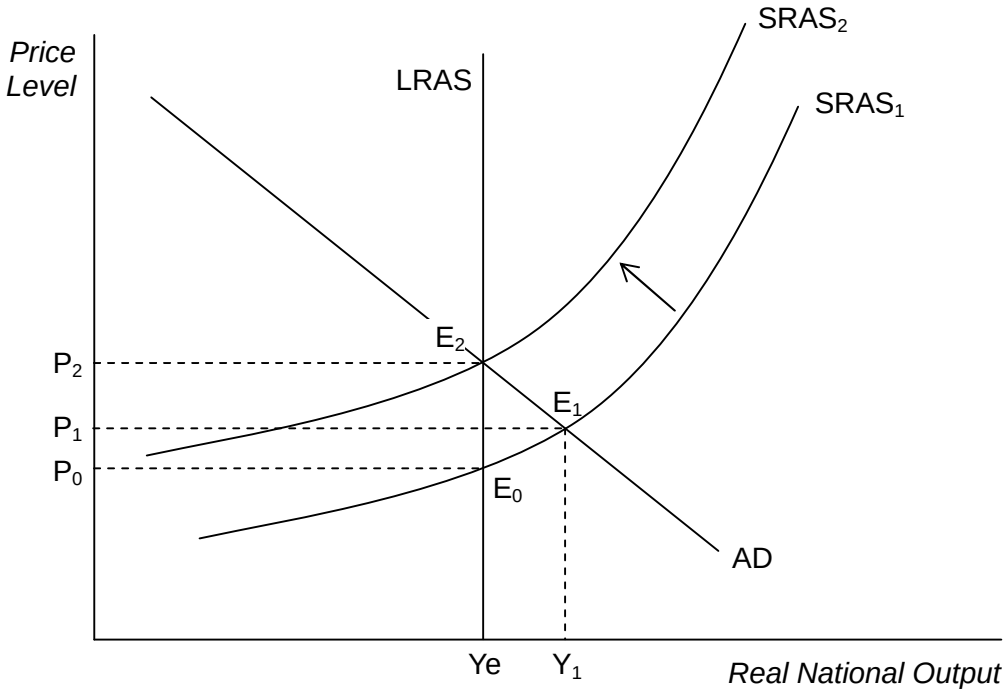


Figure 11.9: Excess aggregate demand

Figure 11.10 illustrates a situation of deficient demand in the economy which corresponds to that termed a deflationary gap in the earlier analysis. The aggregate demand curve AD_1 intersects the SRAS curve at E_1 and the associated equilibrium level of real national output is Y_1. National income level Y_1 is less than the full employment capacity output level of Ye as a consequence of the deficient level of aggregate demand. However, using the AD/AS model we can see that the term deflationary gap is misleading, because the economy may remain in its deficient demand equilibrium at point E_1 without any change in the general level of prices from P_1.

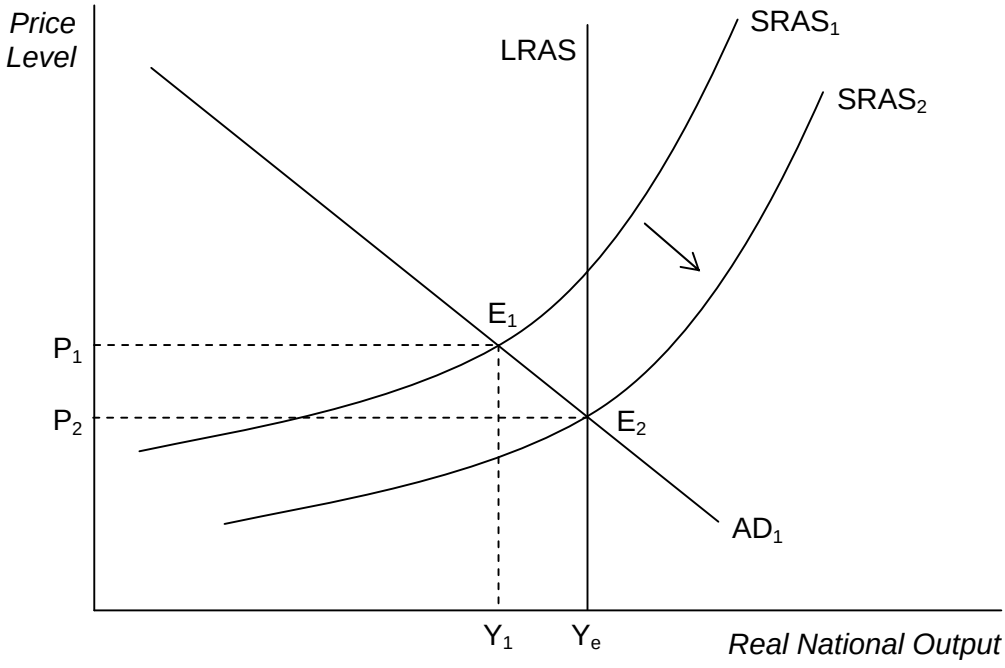


Figure 11.10: Deficient aggregate demand

© Licensed to ABE

What is the significant difference between the situation of excess aggregate demand illustrated in Figure 11.9 and the situation of deficient aggregate demand illustrated in Figure 11.10? In the case of excess demand there a few if any forces in the economy to resist the rise in prices that move the economy to its point of long-run equilibrium. In the case of unemployment due to deficient aggregate demand, the economy's automatic adjustment mechanism will only work if money wages and other costs fall to shift the SRAS curve downwards and to the right, until it intersects the unchanged AD curve at point E_2 on the LRAS curve. If workers resist the attempt to cut their money wages, and it may be individually rational for them to do so, this will prevent the economy from achieving full employment. This is an example of how perfectly rational behaviour on the part of each individual nevertheless leads to a collective or aggregate outcome that is socially undesirable. In this situation, the appropriate policy response by the government is an expansionary fiscal policy to boost aggregate demand, rather than a process of falling wages and prices (deflation), in the economy.

Using Fiscal Policy to Correct a Deficiency of Aggregate Demand

While the concepts of inflationary and deflationary gaps are useful in illustrating the crucial role of aggregate demand in determining the economy's equilibrium level of real output, and hence employment, the neglect of the supply side of the economy fails to reveal the full inflationary consequences of fiscal policy. As explained previously, once the economy is operating on its long-run aggregate supply curve, any additional increases in aggregate demand will merely serve to drive up prices and add to the rate of inflation. But what the analysis also reveals is that even in situations of deficient aggregate demand and unemployed resources in the economy, an increase in aggregate demand will lead to a higher price level and inflation as well as increased real national income. That is, the concept of a deflationary gap for states of the economy involving deficient aggregate demand is misleading, if it is taken to imply that such a state is associated with falling prices.

Figure 11.11 illustrates how using fiscal policy to increase aggregate demand, even when the economy is suffering from deficient aggregate demand, leads to a higher price level as well as an increase in real national output.

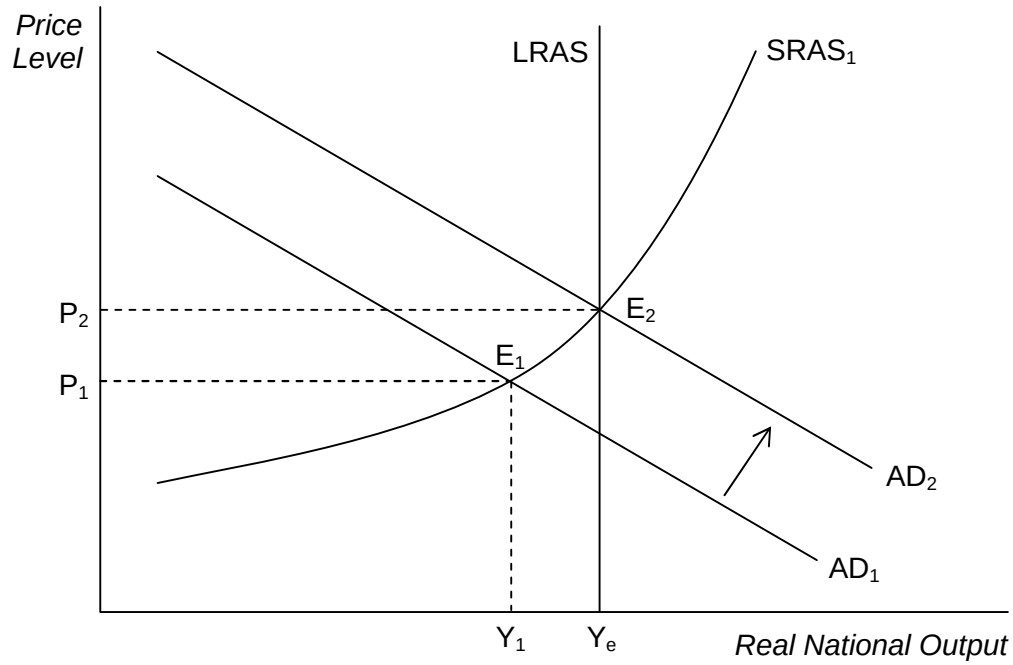


Figure 11.11: Using fiscal policy to increase demand

© Licensed to ABE

The initial level of aggregate demand is shown by AD_1. The initial equilibrium in the economy is at point E_1 where AD_1 intersects with the short-run aggregate supply curve, $SRAS_1$. At equilibrium point E_1 the economy is operating below its full capacity as represented by the position of the long-run aggregate supply curve, $LRAS_1$, at Ye. The economy is suffering from a deficiency of aggregate demand and its shortfall in real output is equal to the distance Y_1-Ye. At the initial equilibrium level of real national output of Y_1 the general level of prices in the economy is P_1. If the government increases its level of expenditure by running a budget deficit to increase the level of aggregate demand in the economy, the AD curve will shift to the right. This is shown in Figure 11.11 by the movement to AD_2. Provided the government's expansionary fiscal policy, which will boost to demand through the multiplier effect, is calculated correctly, the level of aggregate demand will increase until it intersects $SRAS_1$ at point E_2 on the long-run aggregate supply curve. At point E_2 the economy has reached its full capacity point and unemployment will have fallen to its "natural" level. However, in contrast to the earlier 45° analysis of the deflationary gap, the elimination of demand deficient unemployment in the economy has resulted in a rise in the general level of prices from P_1 to P_2, and a rate of inflation calculated as $(P_2 - P_1)/P_1$ per cent. This can be seen by comparing Figures 11.10 and 11.11. In both diagrams the initial point of equilibrium is one involving deficient aggregate demand at Y_1. Without government action to boost AD, as illustrated in figure 11.10, full employment can only be restored by a reduction in money wages and prices that shifts the SRAS curve downwards. Comparing Figures 11.10 and 11.11, the point of full employment equilibrium (Ye) is achieved in both cases, but with the significant difference that the level of prices in the economy is higher when aggregate demand is increased through government policy.

Fiscal policy can be used to control inflation if aggregate demand is excessive, but its use gives rise to inflation even when demand is deficient. This is not the only limitation on the use of fiscal policy because, depending upon how it is financed and the economy's exchange rate system, its power to reduce unemployment may be much less than suggested by the kind of analysis shown in Figure 11.11.

F. FINANCING FISCAL POLICY: BUDGET DEFICITS AND PUBLIC SECTOR BORROWING

The difference between a government's total revenue or income and its total expenditure is referred to as its *budget deficit*. In most countries a government's income is mainly obtained from tax revenue. If a government plans to spend more than it expects to receive in the form of tax revenue, or if tax receipts turn out to be less than anticipated when it planned its expenditure, it will have a budget deficit. To fund a budget deficit a government must resort to borrowing. Usually, such borrowing is based on the issue of securities (called bonds) to investors in the capital market. In some countries the main investors are the country's commercial banks. High levels of government borrowing are regarded as bad for an economy, because they can crowd out private business investment and cause inflation. A government budget deficit can cause inflation when a government does not fund its borrowing needs by the issue of bonds to investors, but sells them to the central bank that pays for them by printing money. Governments, unlike you or I, can "borrow" from themselves by legally printing money!

The *public sector borrowing requirement* (PSBR) is the term once used in the UK to describe the government's budget deficit. The term has been given a new name to avoid confusion with the government's net borrowing position. The budget deficit is now termed the *public sector net cash requirement* (PSNCR). The technical terms PSBR and PSNCR are relevant for understanding official government statements and the national accounts, but in everyday usage the term budget deficit conveys the same meaning. For the purpose of simple economic analysis, all three terms can be treated as equivalent.

© Licensed to ABE

By the 1970s the British government was recognising that there was a growing resistance from all sections of the population to high taxation. At the same time there was strong resistance to reductions in what was regarded as socially desirable public sector expenditure. There was an evident temptation for the government to evade its difficulties by the short-term remedy of increasing its borrowing. The monetarist inclined governments of the 1980s sought to achieve balanced budgets and limit expenditure to the constraints of its revenue receipts. However its apparent success in this objective was achieved by the device of privatisation. This brought the government large sums of capital which were treated – and spent – as revenue. By the early 1990s there was little left to privatise, and the public sector borrowing requirement (PSBR) again became a major economic issue.

Financing of the PSBR

There are three main sources of finance for government borrowing. These are the *non-bank, non-building society private sector*, the *banks and building societies*, and the *overseas sector.*

The financial instruments whereby the government borrows from these three sources are all roughly the same, although of course their relative importance is different in each sector. The main instruments are:

- *Notes and coin* – the cash we carry in our pockets is technically considered to be finance lent to the government. This dates from the origins of the bank note as a receipt of money deposited with a bank. Although we no longer have to deposit gold or silver with the Bank of England to obtain a Bank of England note, this still takes the form of a receipt. Bank notes continue to carry the (now meaningless) "promise to pay the bearer on demand the sum of ... ". The government could increase its borrowing by ordering the central bank to print more and more notes. If it did so the notes would soon lose their value and acceptability.
- *Treasury bills* – these are a kind of very formal IOU, issued for large sums and sold to banks and other institutions prepared to lend money to the government on a short-term basis.
- *Bonds* – bonds known as "gilts" (gilt-edged securities) are issued by the Treasury to banks and other financial institutions as well as the public. Once they have been issued they are marketable, i.e. they can be bought and sold through the Stock Exchange at their current market price.
- *Other Government "paper"* – bonds, certificates, and other financial instruments sold directly to the public and not to banks and building societies. The best known of these paper securities are the national savings certificates and bonds that are issued through post offices.

During the 1980s the British government sought to finance as much as possible of its PSBR through the non-bank, non-building society private sector and the overseas sectors. This was because these sources were thought to have less effect on the money supply than borrowing from the banks. However, in a highly complex financial structure such as that of the United Kingdom, there is some doubt as to whether that is really the case. The government was also able to increase its revenue income by privatisation, i.e. by the sale of shares in the former public corporations such as British Telecom, British Gas and British Airways. These were turned into public liability companies and technically transferred from the public to the private sectors of the economy. In doing this the government was accused of "selling the family silver", and there is certainly some doubt as to the long-term desirability of treating as revenue the proceeds of the sale of capital assets.

© Licensed to ABE

The General Government Financial Deficit

The danger with all the major economic indicators is that governments and others find ways to distort them, so that they cease to be a reliable guide to the true position they are supposed to indicate. Some economists argue that the British PSBR can be subject to distortions of the kind produced by privatisation receipts in the 1980s and early 1990s, and is therefore not always a true indication of the relationship between the government's main taxation revenues and expenditure. Likewise, the UK's New Labour governments after 1997 resorted to numerous dubious national income accounting changes. The reclassification of certain categories of government expenditure created the impression that the government's budget deficit, and accumulated debt as a percentage of GDP, appeared smaller than the true economic position. For this reason economists argue that a more realistic picture of the relationship between government revenue and expenditure is provided by the *general government financial deficit* (GGFD). This is a simple measure of the difference between total tax collections and the net spending by the whole of central and local government. It is the measure that the finance ministries of all the member countries of the European Union use to measure the performance of national fiscal policies.

Importance of Public Sector Borrowing

In the short run the amount of savings in the economy is fixed. If the total demand for finance exceeds its supply from savings, the would-be borrowers have to compete for their share of the available supply. Interest rates are the price of money and like any price they depend fundamentally on the interaction of supply and demand.

Consequently, if the government wishes or is forced to increase its borrowing, it has to compete with the business and personal sector. Thus there is a danger that interest rates will rise, even though for other reasons the government might be seeking to keep them low. If the government wishes to borrow from foreign investors, it will have to offer interest rates that are attractive in world finance markets. If there is a fear of inflation in the home economy, the government will have to offer interest rates that are higher than rates applying in countries where inflation is less of a problem. Investors, quite naturally, wish to protect the purchasing power of the money they invest. The level of public sector borrowing is thus one of the factors influencing the level of interest rates within a country.

It is possible that public sector borrowing will increase the money supply and thus contribute to inflationary pressures in the economy. The precise effect on money supply depends on how the money is borrowed. Some economists argue that an increase in public sector borrowing will only increase money supply if the government borrows from banks or building societies. In these cases the government borrowing creates bank assets. These are then balanced by increased lending by the banks, and the money multiplier operates to increase the total of bank deposits within the economy. Bank deposits are the main element in the total money supply.

It can be argued that increased borrowing from the personal, non-banking sector does not have this effect. When the government borrows from private individuals there is simply a transfer of purchasing power from the private to the public sector. The individual cannot spend money lent to the government. There is no direct increase in the amount of money or bank deposits in the community.

This is the direct effect, but indirectly the increased government borrowing may have further consequences which do affect the money supply. If private individuals lend money to the government, they cannot lend the same money to business firms or building societies. These institutions may turn instead to banks for their finance, having been crowded out of personal lending by the government. If business firms have to borrow more from banks because they cannot raise money on the capital market, there will be an increase in bank

© Licensed to ABE

deposits and lending, i.e. an increase in money supply with its potential for increasing inflation.

Borrowing from overseas investors does not increase the domestic money supply, but it does increase expenditure demand. If there is spare capacity in the economy this will increase the demand for resources, stimulate production and reduce unemployment, assuming that the government is going to spend the money borrowed on home produced goods and services. If there are inflationary pressures in the economy, the increased demand may increase these and contribute to rising prices. If the government spends on foreign goods and services it will reduce the credit balance or increase the deficit on the current balance of payments.

It is clear that there will be important economic consequences of a change in government borrowing. What these are depends on the sources of borrowing and on how and where the borrowed finance is spent.

G. THE LIMITATIONS OF FISCAL POLICY

Looking back from the vantage point of the early twenty-first century, it is now obvious that much of the responsibility for the high inflation of the 1960s and 1970s (and with it the eventual increase in unemployment) was due to the mistaken belief that fiscal policy was all powerful, and that governments could use fiscal policy alone to permanently manage the level of economic activity in the economy. The reality is that government expenditure financed by printing money can only achieve one thing if pushed too far: accelerating inflation and rising unemployment!

The policy solutions developed in the 1990s involved recognition of the limitations of the role of the government and fiscal policy in a modern dynamic economy. The success of the new policies is based upon:

- Recognition that fiscal policy cannot be considered independently from monetary policy. The level of government expenditure, and the size of a government's budget deficit in relation to the level of national income, both have serious implications for the money supply and the level of interest rates in the economy. If inflation is to be avoided, the government's borrowing requirement must be financed out of genuine borrowing from the economy, and/or the rest of the world. It must not be financed from the government borrowing from itself by requiring the central bank to print more money for the government to spend. But even genuine borrowing has implications and its own limitations. The more a government borrows, and the greater its budget deficit as a percentage of national income, the more such borrowing pushes up the level of interest rates in the economy. This leads to "interest rate crowding out".

  Increasing levels of government borrowing from savers in the economy through the financial system takes funds away from companies, and the level of private sector investment in the economy is reduced. That is, the value of the government expenditure multiplier is smaller than suggested by the Keynesian model of income determination. This is because it ignores the way that increased government expenditure crowds out private expenditure through its role in pushing up the level of interest rates in the economy. This problem is made worse if the increased government expenditure is spent on increased current consumption rather than increased investment. By crowding out private sector investment the economy's future productivity capacity is reduced, and with it the future growth rate of national income. The end result of increased government expenditure (although it may appear beneficial today, especially if it allows people to increase their consumption) is thus to reduce the growth of future government tax revenue because this will decline with future income! The correct policy response is to avoid interest rate crowding out by

© Licensed to ABE

reducing the need for government borrowing. Government borrowing as a percentage of national income can be reduced in two ways: by increasing taxation as a percentage of national income, so that government expenditure is paid for without the need to borrow; or alternatively, by reducing the share of government expenditure in national income.

- It is also now recognised that increasing the level of taxation in an economy through higher rates of tax, especially taxes on income and company profits, can also be damaging to the economic performance of an economy. Therefore it follows that increasing taxation is not likely to be a long-term solution.

  As we have seen in this unit, the equilibrium level of national output is determined by aggregate demand and aggregate supply. Any analysis of the role of government expenditure in the economy that ignores the affect of fiscal policy on aggregate supply is likely to seriously overstate the longer-term benefits of fiscal policy. Equal attention needs to be given to aggregate supply as well as aggregate demand in the formulation and implementation of fiscal policy. This is done through the government developing *"supply side policies"*. Such policies recognise the negative incentive effects of high rates of income tax on people's willingness to work, and high rates of tax on companies' willingness to take business risks and undertake investment in new productive capacity. Supply side policies aim to reduce the disincentive effects of taxation. They do this by restraining government expenditure and reducing the share of government in national output, stimulating productivity by improving the quality of the labour force through greater emphasis on education and incentives for increased training, and policies to stimulate increased investment in new technologies.
- The share of government expenditure in total national income can be reduced by "rolling-back the limits of the state" through privatising state-owned industries and transferring functions undertaken by government to private sector firms.
- The eventual acceptance by many economists and governments that persistent long-term inflation was essentially due to over expansion of a country's money supply has resulted in a new approach to monetary policy.
- Government expenditure financed by government borrowing from its central bank (printing money) only leads to inflation, and cannot create permanently higher employment and living standards in an economy. Recognition of this has led many governments to adopt a hands-off approach to monetary policy, by transferring responsibility for the determination and operation of monetary policy to their central bank. This is known as *central bank independence* and is usually, but not necessarily, associated with the adoption of an *inflation target* policy by the central bank. What this means is that the central bank undertakes monetary policy, without interference from the government, with the aim of achieving an announced target rate of inflation in the economy. (This is examined further in later units.)
- The final limiting factor with regards to the effectiveness of fiscal policy is an economy's exchange rate system. Fiscal policy is at its most effective in an economy which maintains a fixed value of its currency against another major currency, such as the US dollar or the European Union (EU) euro. The downside of this policy is that it leaves a country open to importing inflation to the rest of the world. This means that those countries that want to achieve a low and predictable rate of inflation (because this is now thought to be the most effective way of supporting economic growth and full employment) must have a freely floating and not a fixed exchange rate. This explains why those countries which have given their central bank its independence from government in the conduct of monetary policy, such as the UK and the EU eurozone countries, allow their currencies to float on the foreign exchange market. It also explains why many economists and governments believed that fiscal policy was more powerful than it proved to be from the 1970s onwards. In the period 1946–1973

© Licensed to ABE

most countries operated a fixed exchange rate for their currency against the US dollar. This was the period of the International Monetary Fund (IMF) fixed exchange rate system. This system finally collapsed in 1973, when the world's leading economies abandoned fixed exchange rates in favour of floating exchange rates because they wanted to control inflation. Fiscal policy is weak in a country which operates with a floating rather than a fixed exchange rate. (This is examined further in later units.)

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. Outline the aggregate demand and supply model of income determination.
2. What is the difference between short-run and long-run aggregate supply?
3. Explain what is meant by a deflationary gap using the aggregate demand and supply model of income determination.
4. Explain what is meant by an inflationary gap using the aggregate demand and supply model of income determination.
5. Explain what is meant by "interest rate crowding out".
6. What is supply side policy? How does it differ from fiscal policy?
7. In an economy operating with a freely floating exchange rate is fiscal policy stronger or weaker than if the economy operated a fixed exchange rate?

© Licensed to ABE

Study Unit 12

Money and the Financial System

Contents ***Page***

A. Money in the Modern Economy **218**
Features and Types of Money 218
Functions of Money 219
High-Powered Money 220

B. The Financial System **220**
Structure of the Financial System 220
The Retail Banks 221
Foreign Banks 222
Money Markets 222
Building Societies 223
Unit Trusts and Investment Trusts 223
Hedge Funds and Private Equity Funds 223

C. The Banking System and the Supply of Money **224**
Money and Bank Credit 224
Credit Creation 224
Illustration 224
The Bank Credit Multiplier 225

D. The Central Bank **226**
The Functions of the Central Bank 226
Modern Central Banking 227

E. Interest Rates **228**
Importance of Interest Rates 228
The Determination of Interest Rates 229
The Pattern of Interest Rates 231

© ABE and RRC

Objectives

The aim of this unit, in conjunction with Study Unit 13, is to explain and evaluate the effectiveness of monetary policy in a closed and open economy and discuss the possible impact of monetary policy on business decision-making.

When you have completed this study unit *and Study Unit 13* you will be able to:

- demonstrate an understanding of the relationship between the banking system and the creation of money
- identify the components of the high-powered money stock and explain why these have a magnified impact on the money supply
- explain the quantity theory of money and its role in explaining the rate of inflation
- discuss the components of monetary policy and explain how they work
- evaluate the factors that determine the effectiveness of monetary policy
- compare and contrast the relative effectiveness of fiscal and monetary policy.

A. MONEY IN THE MODERN ECONOMY

Features and Types of Money

Throughout history money has taken many forms. Almost anything can serve as money as long as people are prepared to accept it in exchange. Acceptability is the one quality that money must have. If this is lost, if people are no longer willing to trust it and thus refuse to take it in exchange for real goods and services, then it is useless.

Other qualities can add to its usefulness. Ideally money should be:

- *portable* – it will not be much use as an aid to transfer if it cannot easily be moved
- *divisible* – it must be capable of reflecting a range of values; animals were once a symbol of wealth but as money they had limitations – a valuation of one and a quarter cows could prove difficult to pay!
- *durable* – saving presents problems if the money saved is likely to die, rot or rust away
- *controllable* – preferably in short supply, not too easily obtained and capable of being controlled by an accepted authority
- *recognisable* – if people cannot recognise money as money they are unlikely to accept it very readily.

One of the oldest forms of money, and one that is still in limited use, is *gold*. When, from time to time, the world economy becomes unstable and other forms of money become less acceptable, the price of gold always rises as people turn (or return) to it as a haven for their threatened savings.

Other precious metals have often been used, especially silver, but this lacks some of the qualities of gold. Many metals suffer deterioration over time.

To aid recognition, add acceptability and assist in measuring value, many communities over the ages have fashioned coins from previous, semi-precious and base metals. With the exception of a limited supply of gold, these are now used mainly for units of low value.

Metal is bulky and expensive to transport in large quantities so, from very early times, traders have used *paper* as a more convenient substitute. Paper has always been used in two ways as money:

© ABE and RRC

(a) As a *receipt or representation of precious metal* or some more solid form of money and exchangeable for the preferred form of money under certain conditions. The Bank of England bank note still contains the "promise to pay the bearer on demand the sum of ... ". At one time the holder could exchange such notes for gold. Today handing over a note at the Bank of England will only be met with another note, but the promise serves as a reminder that the paper really just represents money and has no intrinsic value in itself.

(b) As an *instruction* to a clearly identified person or organisation, or a promise from a person or organisation, to make a payment under certain conditions. A *letter of credit* is an instruction to make money available to the holder while a *bill of exchange*, still widely used in international trade, is an unconditional promise to make a payment. Such instruments of payment are almost as old as trade itself.

In recent years *plastic cards* have replaced or supplemented paper as conveyors of instructions to make payments. The development of modern telecommunications has made such cards, with their magnetic strips, among the most important means of carrying out everyday trading transactions. As information technology continues to advance we can expect these cards to gain further uses, but we can also expect that transactions will be increasingly made by direct instructions through computers or over the telephone.

All of these convenient forms of payment by simple instruction depend on people's willingness to hold their store of money in banks. Early banks actually did store the wealth of their customers in the form of precious metals, but wealth is now stored purely in the form of credit balances recorded in computers. No doubt today's method of storing money has not yet reached its ultimate form, though in simple terms we can ignore all present and future methods of transferring and storing money and simply refer to it as "bank credit". In this form we can choose to store it as a bank deposit or use it to make payments by any of the techniques made available to us by current technology.

Functions of Money

The functions of money are generally summarised as follows:

(a) Facilitating Exchange

The basic purpose of money, as we have already noted, is to make the exchange of goods or services easier. Without money, people have to resort to direct exchange or barter, and this is often wasteful, time-consuming and inefficient. Money allows trade to develop much more freely.

(b) Measure of Value

Even if people do exchange goods directly, they can be more certain of fair dealing if they can measure the value of their goods in terms of recognised money. If farmers wish to exchange pigs and cows, they are helped if they know the values of both in money terms.

(c) Measure of Deferred Payments

Exchange and trade can flow more freely if it is possible to carry forward debts of a known amount. Money can help by standing as a measure for any payments that are deferred for future settlement. For example, the farmers exchanging pigs and cattle may agree that A took cattle from B to a higher value than the pigs he passed to B. If the difference in value is expressed in money, then both know the size of the debt and the future payment required. Money measurement may help them later to settle the debt – say, with some other animal, perhaps sheep.

© ABE and RRC

(d) Store of Value

Finally, money can be kept as a store of value that can be held in reserve for purchases not yet planned. This value can be held over time – as long as money value does not fall.

The importance of acceptability has already been stressed. Without it, money cannot be used in exchange. This is why a great deal of international trade is carried out in a relatively few generally acceptable currencies – e.g. American dollars, Swiss francs, Japanese yen, British pounds and euros. These currencies are all readily acceptable and transferable in world trade and finance markets.

We can see that acceptability and transferability depend on the confidence of traders. If this confidence is lost, then money ceases to have any value, because it cannot fulfil its essential functions.

The function that causes the most problems is that of *storing value*. No form of money in the modern world has escaped the problem of inflation – the tendency for money prices to rise as time goes by. If all prices rise, then the value of money itself is falling. The difficulty of storing value undermines confidence, acceptability and transferability, and so makes trade generally more difficult and uncertain.

High-Powered Money

The measurement of money supply depends on how we define it. The wider our definition, the more we have to measure. Difficulties in deciding precisely what should be counted as money help to account for the fact that there are several possible definitions. These are can be divided into two groups:

- *Narrow money* – M0, the narrowest definition, made up of the notes and coin in circulation with the public and banks' till money and the banks' operational balances with the central bank.
- *Broad money* – M4, made up of notes and coin and all private sector sterling bank and building society deposits.

This distinction is more important than it might appear because of the special role of narrow money in the banking system. The other name for narrow money is *"high-powered money"*. The term "high powered" indicates that it serves as the reserves of the commercial banks in the economy and provides the basis for the creation of bank deposits. Because high-powered money is "created" by the central bank, and hence directly under its control, it enables the central bank to control the deposit creating activities of the commercial banks and the broad money supply.

B. THE FINANCIAL SYSTEM

Structure of the Financial System

The financial system is made up of a range of banking and other financial institutions and financial markets. These have undergone far-reaching changes in many countries in recent years, especially in relation to the development of financial markets. You are likely to find a number of terms used to describe banks and financial markets when you read textbooks and financial journal articles.

The following subsections provide brief outlines of the various categories. You should also be alert for references and descriptive accounts which appear from time to time in the leading financial journals.

© ABE and RRC

The Retail Banks

These are the banks which handle the individual accounts, both small and large, of private and business customers. In the UK they include such banks as Barclays, Royal Bank of Scotland, LloydsTSB, HBOS, HSBC and Abbey. They are distinguished from investment banks, such as Goldman Sachs. Investment banks are major participants in global financial markets and handle only large sums of money (upwards from $1m), and concentrate their activities in a limited number of major world financial centres. The large retail banks (also known sometimes as branch banks) do engage in wholesale banking in addition to their retailing functions, and the terms "retail" and "wholesale" really apply more to functions than to separate, specialised institutions.

The major functions of a retail bank are:

(a) Safe-keeping of Money

This is the basic function of banking. Many customers still keep jewels and important documents in bank safes. However, modern money is mostly in the form of transferable credit, and this function is chiefly performed through the various types of bank account held by customers. The current account is used for day-to-day transactions. Other accounts are usually in the form of "time deposits", i.e. deposits where an agreed period of notice is required for withdrawals without penalty. The longer the period of notice and the higher the amount deposited, then the higher the rate of interest paid by the bank. If immediate withdrawal is required then a certain amount of interest is usually forfeited, though in some accounts immediate withdrawal is permitted without an interest penalty provided a stated minimum sum remains in the account. You should obtain details of the range of accounts offered by your own bank.

(b) Transfer of Money

Much of the daily work of the retail banks is concerned with making payments through cheques, standing orders, direct debits and other written instructions, including bank giro. Some of the work of money transfer has now been passed to the credit card companies (themselves mostly owned by the large banks), but credit card payments still require final settlement by a bank transfer. The large international banks are deeply involved in foreign payments for the import/export trade. Bills of exchange are still used extensively in handling trade payments, especially as these are very closely linked with the extension of credit.

(c) Lending Money

Banks make most of their profits from lending money. Traditionally they have been chiefly concerned with short-term loans – very "short-call" (overnight or 24-hour) loans to other banking institutions, overdrafts, trade loans made by discounting bills of exchange (usually for up to 60 to 90 days) and commercial loans for up to around two years for business or approved private projects.

In recent years, banks have been encouraged (by government pressure or by competition) to lengthen their lending terms. Clearing banks have entered the private house mortgage market where loans can be made for 20 or more years. Of greater importance to business has been the increased willingness of banks to lend for periods of between five and ten years for business capital development.

(d) Money Management, Advisory and Agency Services

The banks have become increasingly involved in selling their financial skills to help people manage their money. They also recognise that they have a responsibility to provide financial help to business ventures which operate with bank money. Apart from becoming financial consultants, banks are also becoming more actively involved in the

© ABE and RRC

fringe financial services such as insurance broking, investment advice and the handling of trusts and estates.

More recently, a number of banks have entered the field of stockbroking. This has been made possible by the Stock Exchange reforms of October 1986. The retail banks also control a number of specialised subsidiaries, offering hire purchase, leasing and factoring services to customers.

- *Leasing* is an alternative to hire purchase, and is used frequently by business firms to obtain vehicles and equipment under a form of instalment credit.
- *Factoring* is used chiefly in foreign trade. A factor takes over responsibility for a company's approved trade debts (debts owed to the company) and arranges collection and administration, thus releasing cash to the company. It is an expensive way of speeding up a firm's *cash flow* (the speed at which money spent on production is recovered from sales) but worthwhile if the cash can be used at greater profit than the cost of the factoring service.

Foreign Banks

A feature of recent years has been the globalisation of banking and financial markets and the continued rise in importance of a number of international financial centres including London, New York, Tokyo, Hong Kong and Singapore. Such centres attract foreign banks and this is especially true of London, which is home to several hundred foreign banks as well as the UK's retail banks. On the whole, there has not been any major or sustained competition for the business of British industrial companies. Most foreign banks are concerned chiefly with their own national organisations and with operations in wholesale banking – i.e. lending large sums to other banks and financial institutions, usually on a short-term basis. The increase in oil wealth has encouraged the entry to London of a number of Middle Eastern banks.

The foreign banks are also active in what is termed the eurocurrency market, which handles transactions in the bank deposits of banks held outside the banks' countries of origin. Thus the dollar deposits of an American bank in London form part of the eurodollar market in Britain. Eurocurrency markets have become a major part of the wholesale banking structure.

Money Markets

The term "money markets" is given to the markets in short-term money, in which all the main banks, domestic as well as foreign, investment as well as retail, take part. By short-term (when describing money markets) is meant a period of time from 1 to 364 days. Transactions in funds for periods of a year or longer are usually termed capital market transactions, to distinguish them from the very short-term nature of transactions in the money markets, most of which are for days or weeks rather than months. There are a number of different money markets in a developed financial system such as that found in the UK, the EU and the USA. The most important money markets in the UK are the gilt repo market (sale of gilt-edged securities), the interbank market, the certificate of deposit (CD) market, and the commercial paper (CP) market.

These markets bring together domestic and foreign business organisations, all banks as well as central and local government, all of which have funds that they have to keep almost liquid but which they cannot afford to have lying idle. In the money market funds are not allowed to lie idle. When London sleeps its money may be working hard in Sydney, Hong Kong, Singapore, Tokyo and many other places. If you have £10 spare you will not earn much interest by lending it overnight, but if you have £10 million it could easily be earning over £1,000 while you sleep – and still be back in your account next morning ready to meet any payment due to be made.

© ABE and RRC

Building Societies

Historically the main function of these institutions was the provision of funds for house purchase by individual owner-occupiers. They are also a major channel for the savings of individuals. The societies have expanded with the huge growth of private home ownership in the United Kingdom. At the same time, there have been many mergers so that the number of societies has been falling, but their average size has increased. The *Building Societies Act 1986* opened the way for the larger building societies to convert to public companies as well as becoming full banks.

Life Assurance Companies and Pension Funds

These are the most important financial institutions in terms of their role as the main long-term investing institutions in the economy. The life and pension companies differ from general insurance companies in that they provide long-term investment services, and do not normally sell protection on an annual basis. For instance, a payment made for motor insurance covers the cost of protection for the year of insurance. The premium thus buys a specific and limited service. The typical life assurance or pension contract provides for a return payment to be made at some time in the future, prior to which there is a continuing obligation to pay premiums and a continuing obligation on the part of the company to invest those premiums to the mutual benefit of the company and its policy holders. This gives the life and pension companies substantial funds which they invest in a range of ways including property, shares, and government bonds – or in direct lending to business. Today in the UK they are the main investors and holders of company shares, corporate and government bonds, and major participants in the financial markets.

Unit Trusts and Investment Trusts

These represent slightly different forms of pooling revenues to spread the risks of investment.

Unit trusts are the more popular. A trust sets up a fund which is invested in a published range of securities. The fund is divided into units of fairly small denominations which are then sold to savers in a variety of ways. The unit trust holder thus has his or her savings effectively spread over all the funds' investments. Units are bought and sold by the managers of the fund, so that they do not pass through the Stock Exchange. The managers of course deal through the Stock Exchange in the course of managing the fund's investments.

Investment trusts are limited companies which use their share capital to invest in other companies. Their own shares are bought and sold through the Stock Exchange, and shareholders are effectively investors in a range of other shares.

Hedge Funds and Private Equity Funds

In addition to unit and investment trusts (traditional examples of collective investment organisations), there are a wide range of mutual funds and other more specialised forms of investing institutions. These include hedge funds and private equity funds. Hedge funds originated in the 1950s but have only risen to importance (and made news headlines) since the 1990s. Hedge funds are intended for very wealthy investors, rather than the average retail saver who invests through life insurance, unit trusts and mutual funds. Hedge funds employ a wide range of strategies to try and achieve high rates of return irrespective of the state of the economy. However, the fundamental rule in finance is that consistently high returns on investments are impossible without taking on very high risks! What this means is that while some hedge funds do produce high returns, others make equally spectacular losses, which is why they are restricted to very rich investors not small savers.

© ABE and RRC

C. THE BANKING SYSTEM AND THE SUPPLY OF MONEY

Money and Bank Credit

Disagreements between economists about the motives for holding liquid money in preference to other forms of wealth may not seem too important. In practice, they affect government economic policies and the way they seek to control the economy through interest rates. Anyone with a house mortgage or a bank loan knows only too well the effect of changes in interest rates.

In order to take our understanding of the issues a little further, we must examine the relationship between the demand for money and its supply.

The notion of a relationship between demand and supply may surprise you. In our earlier examination of demand and supply for goods and services, these two market forces were kept separate. However money is rather different. It is not "produced" like other commodities, except in the very limited sense that gold and silver are produced. As we have seen, most of the supply of modern money is not found in physical form at all – it is *credit* held in bank accounts on behalf of the banks' customers. The total amount of credit held by the banks on behalf of customers is not a fixed amount; it can itself be varied by the banks' own actions.

Credit Creation

In fact banks can create credit through lending to their customers, and lending is a most important – and profitable – part of a bank's activities. When people or firms borrow from the banks, they use the amount borrowed to make payments to other people or firms, who deposit the payments with their own banks. Suppose I borrow £2,000 from my bank to help buy a new car. When I buy the car, I pay the Swifta Motor Company. Suppose this company also has its accounts at the same bank. When I pay my cheque (drawn on the bank) to Swifta, it then pays in my cheque to its own account. In effect, the bank has created £2,000 in one account (my loan account) and thereby increased the volume of its customer deposits (through the extra £2,000 paid in by Swifta).

Thus, for the factor capital, we have the peculiar position that demand appears to create its own supply.

You may think we have cheated by using one bank only in our example but, as long as there is a fairly closed banking system in a country, the effect will be the same if different banks are involved. In the UK, the great mass (over 80 per cent) of daily payments pass between the four large clearing banks (Barclays, LloydsTSB, NatWest and HSBC), so that this close relationship between demand, borrowing, depositing and supply does exist.

Illustration

In practice, the banks keep a proportion of all their funds in the form of coin, notes or deposits with their own bank (the Bank of England), or in loans to other banking institutions, which can very quickly be recalled. Such funds are the cash reserves of a bank and referred to as high-powered money. If we call these reserves "cash" and assume, for simplicity, that a country has a system of two banks only, each keeping 10 per cent of its assets in cash, then we can give a very simple illustration of how the total supply of bank money can grow following the injection of "new money" from some outside source.

Suppose that our two banks are A and B, and the initial injection is 100 currency units, which goes to bank B. Bank A's customers borrow money to pay to customers of B, and vice versa. The banks are of equal size.

© ABE and RRC

Bank A

Customer deposits	1,000	*Held as:*	Cash	100
			Loans	900
				1,000

Bank B is in the same position. Then there is an injection of 100 to the deposits of A. Bank A initially adds this to its cash – but idle cash earns no money. Therefore as soon as possible it lends it to suitable customers, and its accounts then appear as follows.

Bank A

Customer deposits	1,100	*Held as:*	Cash	110
			Loans	990
				1,100

This additional lending soon gets paid into customer deposits of bank B, which also lends 90 per cent of this increase, so that its accounts appear as:

Bank B

Customer deposits	1,090	*Held as:*	Cash	109
			Loans	981
				1,090

Additional loans of 81 units have now been made to customers of bank B, who have made payments to customers of bank A.

The process continues, and bank A's accounts become:

Bank A

Customer deposits	1,181	*Held as:*	Cash	118
			Loans	1,063
				1,181

Notice how the total of deposits (and hence the total money supply) is increasing, but (because 10 per cent is being held back all the time) by a decreasing amount at each lending/deposit round.

The Bank Credit Multiplier

This progression is called the bank credit (or money) *multiplier*. The total increase in our example will be ten times the amount of the original injection. This is because:

$$K_b = \frac{1}{c}$$

where:

K_b = value of the bank credit multiplier

c = proportion of customers' deposits held by the bank as "cash".

In our example, the proportion held as cash is 1/10 and so the value of K_b is 10.

© ABE and RRC

As the original injection was 100 (currency units), the final increase would be 1,000. Thus, the greater the proportion of customer deposits that the banks are able to lend to other customers, the greater will be the size of the bank multiplier and the effect of lending on total money supply.

This power of the banks to "create money", and the close link between lending money and the increase in total money supply, are both extremely important issues. You must make sure you fully understand them.

Because of this close relationship between the demand for and the supply of money, we can suggest that the supply of money is likely to have very similar features to the demand. Thus, if we believe that there is a particular relationship between interest rates and the demand for money, then a very similar relationship can be expected for interest rates and the supply of money.

D. THE CENTRAL BANK

Of rather greater economic importance is the *central bank* – in the UK this is the Bank of England. The central bank does not compete for ordinary commercial banking business. It is, essentially, the banker to the rest of the banking system: the regulating body for private sector commercial banking and the office link with other central banks and with international financial organisations, especially the International Monetary Fund (IMF) and the Bank for International Settlements (BIS).

The Functions of the Central Bank

The traditional functions of a central bank (further explanation of which can be found on older textbooks on money and banking) are:

- *Banker to the government* – the government holds its bank account with the central bank. This is used both for payments made from the rest of the economy to the government and payments by the government in the economy. The central bank may also be banker to the government in the sense that it provides loans to the government, as well as arranging for the government to borrow from investors in the financial system by issuing treasury bills (short-term securities) and bonds (long-term securities).
- *Banker to the banking system* – the central bank provides the paper currency and coin issued to the public through the banking system. As banker to the banks, it keeps the accounts of the retail banks themselves. It facilitates the process of *clearing* the daily balances resulting from all the transactions undertaken each day by the customers of the banks when they receive and make payment using their bank accounts. In the UK, the banks that maintain accounts with the Bank of England for the purpose of settling the interbank debits and credits that result from their customers daily cheque transactions are termed "clearing banks".
- *Lender of last resort* – the central bank is uniquely placed to lend to other banks in the financial system because it manages the government's accounts, and can call upon the government to print more money in an emergency situation. The central bank acts as lender of last resort to the banking system in two ways:
    - It controls the available supply of liquidity in the banking system on a daily basis to maintain interest rates at the level it thinks appropriate to achieve its monetary policy objective(s). It does this by determining, on a daily basis, the rate of interest at which it is willing to provide funds to any bank facing a shortage of liquidity, in exchange for government bonds and treasury bills.

© ABE and RRC

- It stands ready to prevent the failure of any bank, and the loss of public confidence in the soundness of the banking system, by providing emergency loans to one or more of the retail banks in the economy. This would be necessary if the banking system as a whole runs short of liquidity due to factors unconnected with the central banks' own monetary policy actions. An example of this is provided by the Bank of England's emergency support for Northern Rock bank in 2007.

- *Regulation and supervision of the banking system* – it is responsible for the stability and integrity of the institutions which make up the banking system.
- *Monetary policy* – it is responsible for the conduct of monetary policy. In the UK the Bank of England has a duty to control the actual supply of money within the banking system. The reasons for monetary controls, and the ways in which they may be exercised, are examined in Study Unit 13.
- *Management of a country's foreign currency reserves* and responsibility for its *exchange rate policy*.

In the UK the Bank of England keeps the nation's gold reserves and the international accounts for money entering and leaving the country, as well as the nation's reserves in other currencies. The Bank of England works closely with the central banks of other nations.

The Bank maintains continuous contacts with the major international banks, especially the International Monetary Fund (the IMF is probably closest to being a genuine world bank).

The Bank has a duty to maintain the stability of the national currency in its exchange value with other national currencies, and to cooperate with other countries and international institutions to uphold the stability of the world financial system. It has a special account which it can use to deal in sterling and other currencies in order to stabilise demand, supply and exchange rates.

Modern Central Banking

Since the 1980s there has been an increasing trend by countries to change the role of the central bank and reduce its functions. This is why the functions just detailed are referred to as the "traditional" functions. The modern trend is to separate the functions of financing the government, regulation and supervision of the banking system and monetary policy. The central bank is given primary responsibility for using monetary policy to achieve a low rate of inflation. The Ministry of Finance or Treasury is given full responsibility for funding government borrowing. A separate financial regulatory authority is given responsibility for the regulation and supervision of the banking system. In the European Union the European Central Bank (ECB) is solely responsible for the formulation and operation of monetary policy, independently of all the EU eurocurrency zone governments. In the UK the Bank of England has operational independence for monetary policy, while the Financial Services Authority (FSA) is responsible for the regulation and supervision of the financial system. The UK Treasury is now solely responsible for managing the national debt and the financing of additional government borrowing from the financial system. The reasons for this development are considered further in Study Unit 13 dealing with Monetary Policy.

© ABE and RRC

E. INTEREST RATES

Importance of Interest Rates

We have seen how important borrowing and lending are to the workings of a modern economy, but this dealing in money always takes place at a price. The price of money is *interest*, and the level of interest has become an important issue in modern economics. The reasons why interest rates have gained this importance include the following:

(a) Interest rates influence the level of business investment and business costs

If interest rates are high, new investment is discouraged. As most loans provide for interest rates to be linked with bank base rates, the costs of existing borrowing rise. The result of a prolonged period of high rates is that business efficiency declines. This reduces the supply of business goods and services, and makes it more difficult for businesses to compete with countries with lower interest rates.

(b) Interest rates influence the cost of public borrowing

The government, in one form or another, is by far the largest borrower of money. Some government debt is subject to changing rates. A number of loans are linked to rates of price increase, and the government's short debts (treasury bills) have to be constantly renewed at current market rates. Governments have to pay interest out of revenue, and taxation is the largest source of revenue. A large proportion of tax revenue thus has to pay for the costs of past spending, and this proportion is not available for new spending. Any rise in interest rates reduces the amount of public services that can be provided from taxation, and makes the government dependent on further borrowing – thus increasing future costs still further.

There is also a social effect. Remember that taxes are paid, directly or indirectly, from income earned by labour. Interest goes to holders of capital, so that the higher the rate of interest, the greater the effective transfer of income from labour to capital.

(c) Interest rates influence consumer spending

Much consumer spending on major capital goods and the more expensive household durables is with the help of credit. If interest rates are high, consumers may go on spending for a time but:

(i) they purchase less expensive goods, because a higher proportion of the amount spent goes on borrowing costs, and

(ii) the burden of repayments takes up an increased proportion of income – leaving less for other spending.

As everyone with a mortgage loan knows only too well, any increase in the interest charged on the loan reduces the amount of household income left for spending on other goods and services. If for any reason the household cannot meet the mortgage payments the home may be repossessed. Changes in the rate of interest have become of very great importance to large numbers of people.

High interest rates also appear to increase savings – partly, no doubt, because of the discouragement to spending. An increase in saving and a reduction in consumer spending can have a depressing effect on total business activity. A prolonged period of very high rates can be an important influence leading to general depression and increased unemployment.

© ABE and RRC

(d) Interest rates affect the rate of inflation

Because interest rates affect the cost of consumer spending, and because building society and bank mortgage interest rates now affect around 60 per cent of all households in Britain, any change in rates influences movements in the Retail Price Index, the official measure of average price increases (inflation). If interest rates go up, then inflation rises and people tend to spend less on new purchases. If spending also falls, then unemployment may rise, even though prices are also rising.

Because of the direct impact of interest changes in all these ways, the ability to make changes has become a major instrument of economic policy in all the main market economies. Since most contemporary governments in the advanced market economies appear to be pursuing mainly monetarist, anti-inflationary policies, they all rely on interest rates to pursue their objectives.

The Determination of Interest Rates

Since interest rates have so many important influences on our lives, we should have some knowledge of the processes which determine them. Interest is of course the price of money, so that ultimately we would expect the forces of supply and demand in the finance markets to determine the levels of interest ruling at any given time. This in fact is the basis for one of the most widely accepted theories of interest rate determination. This theory suggests that the market equilibrium rate of interest is that rate at which the stock of available capital is equal to the demand for capital arising from its marginal efficiency.

The marginal efficiency of capital within the community is the average return from capital investment available to business organisations. Our earlier discussion of business investment showed that business firms can be expected to invest capital and to acquire capital for investment as long as the return from investment is more than the cost of capital. In this analysis, we can equate the cost of capital with the market rate of interest. Firms will not knowingly invest where the return is less than the cost of capital (market rate of interest). The interaction of supply of capital and its marginal efficiency is illustrated in Figure 12.1.

As there is only a limited number of high-yielding investment projects, we can expect the marginal efficiency of capital (MEC) to fall as more capital is invested. The MEC curve is thus downward sloping. The stock of capital is fixed at any given time, and is shown by the vertical line which intersects with the MEC curve at interest rate i and quantity of capital q. At this rate and quantity the demand for capital resulting from its MEC is just equal to its supply – the capital stock – so that demand and supply are in equilibrium at interest rate i. At any higher rate there is an excess of demand as at rate i_1, where demand rises to q_1 with supply remaining at q. At rates above i there would be an excess of supply over demand.

© ABE and RRC

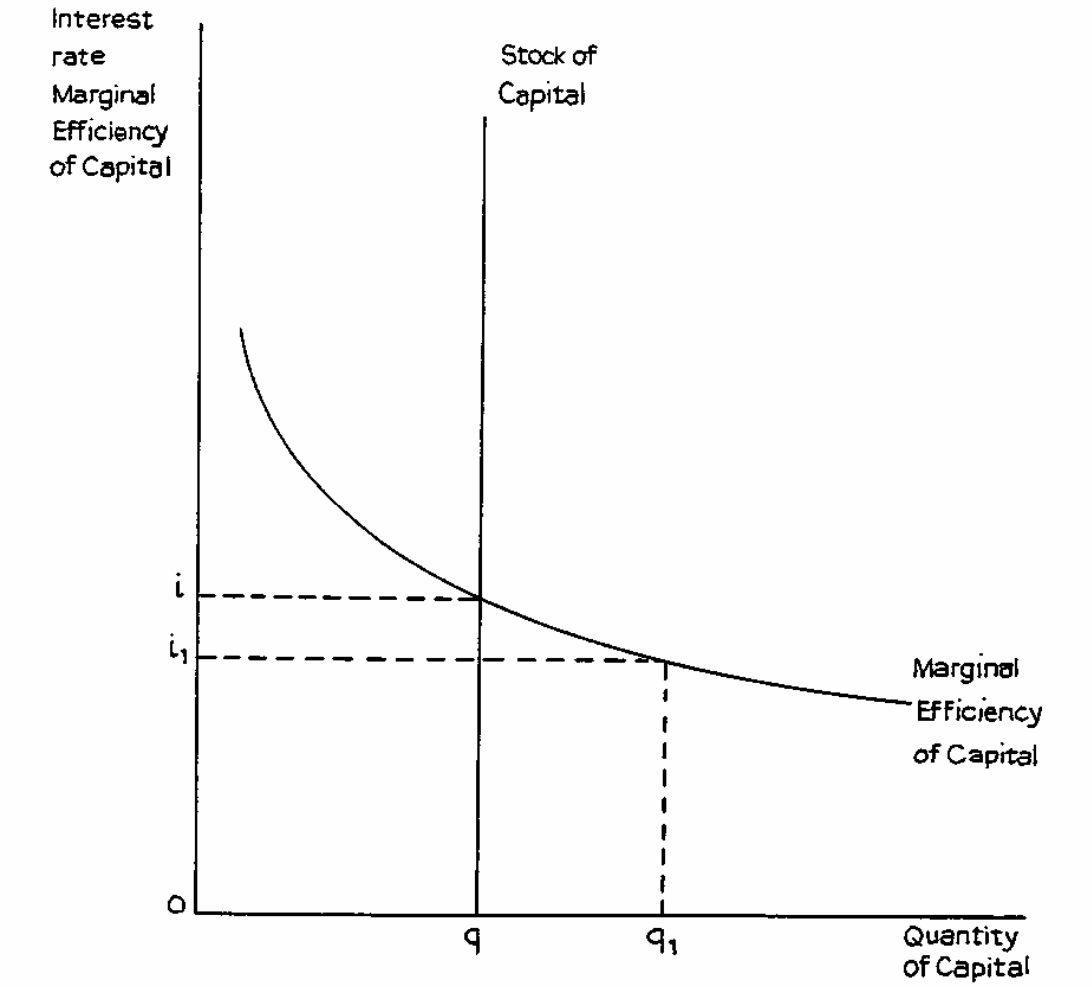


Figure 12.1: The interaction of supply of capital and its marginal efficiency

In the absence of any other influence, interest rates would be determined by considerations of this nature. However, other influences are almost always present in the shape of government or central bank intervention. Because some governments or central banks intervene to move interest rates to levels thought necessary to achieve their desired economic objectives, other governments also have to intervene to ensure that their economies are not put at a disadvantage.

Governments or other regulatory bodies are likely to want to push rates higher than the market equilibrium levels, if they wish to restrict demand and production in order to control inflationary pressures. They may seek to bring rates below the equilibrium if they are faced with high and rising unemployment and fear a deep recession-depression. By reducing the cost of capital they hope to encourage business investment and consumer demand for goods and services. No major trading country can afford to be too far out of line with interest rates in other countries, otherwise there would be a huge movement of capital towards high-rate countries and away from low-rate countries. This movement would put immense strains on the low-rate countries' balance of payments and on their currency exchange rates. Consequently the freedom of any individual government or central bank is restricted by the actions of governments and banks in other countries. Finance now circulates in a genuinely international market.

Governments can influence rates either by controlling the stock of capital, usually by measures over bank lending, or by direct controls over the major banks. Notice that in Figure

© ABE and RRC

12.1 the equilibrium rate will rise if the stock of capital line moves to the left and fall if it moves to the right. This results from the general shape of the MEC curve.

The influence of the demand and supply of money, and the control of interest rates through monetary policy, is examined in Study Unit 13.

The Pattern of Interest Rates

Of course it must not be assumed that the market rate of interest applies to all borrowers and lenders. In the first place financial institutions always charge their borrowers a higher rate than they pay to depositors. In general those who lend money to others require a rate of interest which reflects:

- The time period over which the loan is made. The longer the period the higher the interest rate required, unless market rates are expected to fall over the period, when long-term rates can sometimes fall below those for short-term lending.
- The ease with which money loaned can be recovered: the greater the degree of liquidity. The more speedily and simply the money can be recovered, the lower the rate of interest. Banks pay a higher rate on deposits where several months' notice is required before repayment is made than on deposits which offer "instant access" (immediate cash withdrawal).
- The credit standing of the borrower – large companies with a long record of financial stability can obtain loans at lower rates than new, small companies.
- The degree of risk, which is in fact the underlying factor in all the above considerations. Share dividends are not the same as interest payments but very similar principles apply. If you look at the dividend yield as shown in a share price list in any of the leading daily papers, you will see that the yield (dividend return as a percentage of the price of the share) is much lower on shares in the most profitable and secure companies than on shares of small companies in the riskier sectors of activity, e.g. house builders.

You should examine the deposit accounts offered by several of the main banks and see how far the differences in interest rates offered can be explained by these factors.

© ABE and RRC

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved those learning objectives covered in this unit. If you do not think that you understand these objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. What is the difference between narrow and broad money in an economy?
2. What is high-powered money?
3. What is the bank credit multiplier?
4. Explain, using the bank credit multiplier, how an increase in the amount of cash (high-powered money) in the banking system will affect the value of bank deposits and the broad money supply.
5. What is the marginal efficiency of capital?
6. Explain how a reduction in the level of interest rates can affect the volume of bank lending and the level of investment in the economy.

© ABE and RRC

Study Unit 13

Monetary Policy

Contents ***Page***

A. Options for Holding Wealth 234
Physical Assets 234
Financial Securities 235
Liquid Money – Cash 235

B. Liquidity Preference and the Demand for Money 236

C. Implications of the Interest Sensitivity of the Demand for Money 238
Interest Rates and Demand for Goods and Services 238
Classical and Monetarist View 239
The Keynesian View of Interest Rates and Expenditure 239
Implications of the Differences 239

D. Changes in Liquidity Preference 241

E. The Quantity Theory of Money and the Importance of Money Supply 242
The Money Equation 242
Diagrammatic Representation of the Quantity Theory of Money 242

F. Methods of Controlling the Supply of Money 244
Interest Rate Control 244
Control over Banking Ratios 244
Direct Controls over Banks 244
Control of Government Borrowing 245

G. Monetary Policy and the Control of Inflation 245

© ABE and RRC

Objectives

The aim of this unit, in conjunction with Study Unit 12, is to explain and evaluate the effectiveness of monetary policy in a closed and open economy and discuss the possible impact of monetary policy on business decision-making.

When you have completed this study unit *and Study Unit 12* you will be able to:

- demonstrate an understanding of the relationship between the banking system and the creation of money
- identify the components of the high-powered money stock and explain why these have a magnified impact on the money supply
- explain the quantity theory of money and its role in explaining the rate of inflation
- discuss the components of monetary policy and explain how they work
- evaluate the factors that determine the effectiveness of monetary policy
- compare and contrast the relative effectiveness of fiscal and monetary policy.

A. OPTIONS FOR HOLDING WEALTH

There are three main ways in which wealth may be held. These are generally described as:

- physical assets
- financial assets (securities such as bonds and shares traded on stock exchanges)
- cash (liquid money).

Physical Assets

Examples of physical assets would include houses, land, furniture and private cars. Everyone who has wealth of any kind will have some assets, as these are necessary to everyday life in a modern society, but it is also possible to hold the wealth you wish to store for the future in the form of assets. In this case your choice of which assets to hold will be guided less by what you need or find useful in normal life, but by what you think is most likely to hold or increase its value in the future. Since the future is uncertain you may or may not choose correctly!

Holding wealth in the form of physical assets offers the following advantages:

- They are likely to be useful or enjoyable as well as valuable, and may remain so even if they lose their value; for example, vintage wine may not increase in value as hoped at the time of purchase, but it is very pleasant to drink.
- In periods of inflation or financial/political uncertainty, they are likely to hold or increase their value when money is losing its purchasing power.
- They are visible symbols of wealth and status and this can be important for some people.

On the other hand there are some serious disadvantages:

- They can excite envy and attract thieves; if as a result they have to be stored in a bank vault, they cannot be enjoyed.
- They can be destroyed by fire or accident, or damage may reduce their value.
- Keeping physical assets involves costs such as insurance premiums, maintenance, cleaning and guarding; and these costs can be heavy.

© ABE and RRC

- Fashions change, and what is in demand and valuable one year may be considered unattractive and without value a few years later. This applies particularly to the so-called "collectibles", such as works of art, coins and postage stamps. Those who bought houses in the late 1980s know only too well that asset values can fall as well as rise.

Therefore under normal circumstances, few people with wealth to store are likely to hold all their wealth in the form of physical assets. This would be an option only when the normal financial system was in danger of collapsing.

Financial Securities

Financial securities are mostly either titles to the ownership of property or rights to share in the benefits of property ownership, or they are *promises* to make a future payment. It is often an advantage to hold a written title to property, because ownership can be transferred by handing over the written title or it can be used as a security for a loan. Similarly a written promise to make a future payment will also have a value, and the right to receive the payment can be sold to someone else.

To be useful as a financial instrument of course, the promise to pay must carry respect. An undertaking by a major High Street bank will be more transferable, and therefore useful, than one signed by an unknown individual. Such promises to pay or to repay a loan or debt on or by a stated date, with interest payable to the holder in the meantime, are often known as bonds or stocks. There are several different kinds of bonds issued by borrowers, but the most common have the important feature that they pay a fixed annual rate of interest, (usually referred to as the *"coupon"*) to the investor holding the bond. The main categories of bonds are government bonds and corporate bonds. In the UK bonds issued by the British government are termed "gilt-edged securities" (gilts) and are an important element in the capital market. Details of these can be obtained from most post offices and their market prices are quoted daily in the financial press.

Wealth held in the form of bonds and securities, including the ordinary shares of companies, can also be referred to as *loanable funds*. Besides ease of transfer, holding wealth in this form has the advantage that it provides the holder with an income from interest or dividends paid by the issuer of the securities. This is in contrast with owning physical assets, which incurs costs of maintenance and insurance. As with any form of wealth there are risks of suffering a loss. For example, if a company which has issued bonds fails and goes into liquidation with insufficient assets to meet its obligations to bondholders, then the bonds are worthless. The bonds of very risky companies are frequently called "junk bonds".

Liquid Money – Cash

Liquid money is most likely to be in the form of bank credit held in current accounts which, technically, are "sight deposits", i.e. depositors can withdraw or transfer money without having to give notice to the bank. Most people will hold some liquid money in order to make payments by cheque, plastic card or cash in the form of notes and coin. However, since sight deposits generally earn only insignificant rates of interest, if cash were wanted purely for payment purposes the majority of people would keep only the minimum needed for their regular payment needs. In practice, many people with sufficient wealth to be able to choose between the three options may keep liquid money in preference to assets or securities.

Classical economists offered little explanation for this tendency, since they believed that the desire to hold money in its liquid form depended mainly on the desire to use it for making purchases. They did not attempt to relate the demand for liquidity to any other single variable, such as interest rates. That such a relationship could exist was argued by the great Cambridge economist of the 1930s, John Maynard Keynes, whose view of the elements in the demand for liquidity, i.e. "liquidity preference", we will now look at.

© ABE and RRC

B. LIQUIDITY PREFERENCE AND THE DEMAND FOR MONEY

Keynes believed that there was a connection between money and the level of interest rates in the economy, and in his analysis he concentrated his attention on the choice between holding money (liquidity) and bonds. He identified three elements in the attraction of money.

In doing so, he effectively elevated money to the status of a commodity for which there is a demand in its own right – not simply as something to hold when other forms of wealth are temporarily out of favour. The three elements in the preference for liquidity in Keynes's theory are the transactions, the precautionary and the speculative motives.

(a) Transactions Motive

This is the desire to hold money because it is needed for the purchase of goods and services, i.e. to carry out trading transactions.

(b) The Precautionary Motive

This is the need to have some liquid money available as a precaution against unexpected developments, including favourable opportunities to purchase goods.

(c) The Speculative Motive

It is here that Keynes parted company from earlier teaching. Something of a financial speculator himself, Keynes regarded the speculative element, as in the choice between bonds and money, as particularly significant.

The opportunity for speculation (gambling) arises out of changes in interest rates, and the fact that the interest on bonds is normally paid at a fixed rate. Suppose a bond's fixed interest rate was five per cent – because it had been first issued at a time of fairly low interest rates, when people were content to receive five per cent on their money. Suppose that some years later interest rates in general had risen to 10 per cent, so that anyone lending money at that time would want at least 10 per cent from the borrower. Clearly, anyone holding a five per cent bond would not be able to sell it to another at its original price. A purchaser would expect to receive two £100 bonds for every £100 paid, because only then would he be able to secure a total interest payment of £10, which is the amount he could obtain by lending his £100 elsewhere in the financial marketplace.

Thus, with market rates of interest at around 10 per cent, we could expect the market price of a £100 bond paying fixed interest of 5 per cent to be £50.

Now, suppose the market rate of interest started to fall, so that the best rate a lender could obtain was 7.5 per cent. Anyone willing to buy bonds would now be prepared to pay somewhere around £67. (If you cannot see why, then work out how many £100 bonds, paying interest at 5 per cent per year, you would need to give yourself an annual payment of £15 in return for a total payment for the bonds of £200. When you have decided that, then work out the price per bond.) This means that a fall in interest rates from 10 per cent to 7.5 per cent would enable anyone who had purchased a 5 per cent bond for £50 to sell it for £67 – a handsome profit, especially if the change had taken place over a fairly short time period.

We can deduce from this that, if interest rates are high and expected to fall, people would wish to buy bonds. As bonds and money are seen as alternative forms of holding financial wealth, the demand for money would consequently be low. By the same reasoning, if interest rates are perceived to be low and expected to rise, people would not want to be left holding bonds the value of which, as financial assets, is falling. Instead they would sell bonds and hold money – the demand for which would

© ABE and RRC

thus be high. Roughly equivalent to bonds are ordinary shares of first-class industrial and commercial companies, the profits of which might not be expected to fluctuate greatly and the dividends of which are fairly constant.

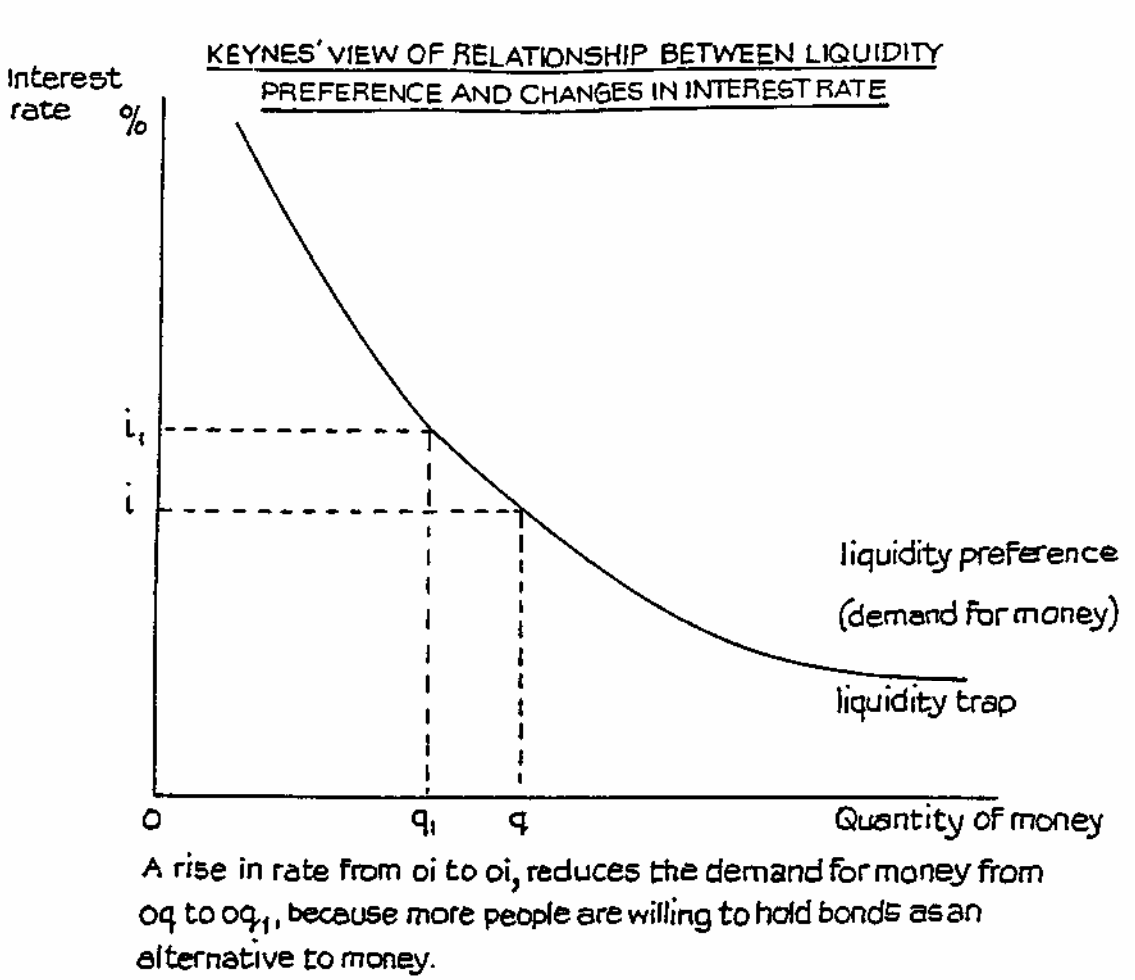


Figure13.1: Liquidity preference curve

What is high and what is low in relation to interest rates depends on a great many other considerations, including people's experiences of rates in recent years. The 10 per cent used in the previous example would have been regarded as very high in the early 1960s, but very low in the early 1980s. You should take an interest in the movement of interest rates and in changes in the prices of bonds (government stocks) while you are studying economics.

This stress on the speculative motive for holding money led Keynes to the belief that the demand for money does have a direct relationship to interest rates. It was thus possible to draw a demand for money curve or "liquidity preference curve" of the type shown in Figure 13.1.

Notice that, at the lower rates of interest, the curve flattens out, creating a so-called "liquidity trap". This is because no one believes that the rate is likely to fall further, so there are no takers for bonds and people will wish to see a rise to a higher rate before there can be any expectation of a fall and a chance for a speculative gain.

The modern view of the influence of money on interest rates gives less emphasis to the speculative demand for money and the idea of a liquidity trap, but rather incorporates the demand for money into the theory of the demand for assets in

© ABE and RRC

general. Modern portfolio theory recognises that there is a demand for money as an asset as well as for transactions purposes, and that changes in the level of interest rates affect the demand for money (see Figure 13.2). However, it is also recognised that there is a very close link between the supply of money and inflation, and that inflation also has a significant influence on the demand for money as well as other assets.

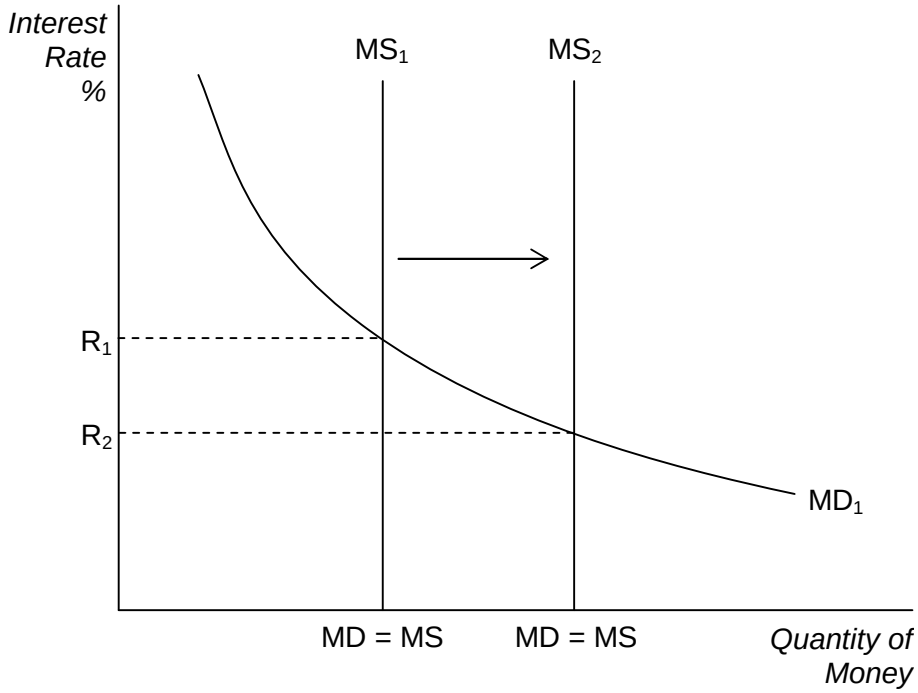


Figure 13.2: Money supply and the rate of interest

C. IMPLICATIONS OF THE INTEREST SENSITIVITY OF THE DEMAND FOR MONEY

Interest Rates and Demand for Goods and Services

We now return to an earlier statement concerning the demand for money. Money is but one of a number of possible ways to hold wealth. Another way is to buy goods, so that we should now consider what is likely to influence the desire to spend money in buying goods in preference not only to holding money, but also to holding bonds or company shares. If we then see interest-bearing or dividend-bearing securities as being in competition with goods for a share of spending, we can also see that bonds, etc., are likely to be desirable, because they yield an income. Goods do not yield an income but they offer other satisfactions. We thus have to balance the desire to obtain an income with the desire to enjoy goods – and services. If interest rates are high, then bonds and other income-yielding securities can seem attractive, because of the income that they produce. If interest rates are low, the income attraction is also low, and goods and services offer greater satisfactions.

Taking this approach, we can see a relationship between movements in interest rates and movements in the demand for goods. When interest rates are high, the demand for goods is low, because people prefer bonds. At low interest rates, demand for goods is high because they seem more attractive than the low income obtainable from bonds. This relationship is shown in Figure 13.3.

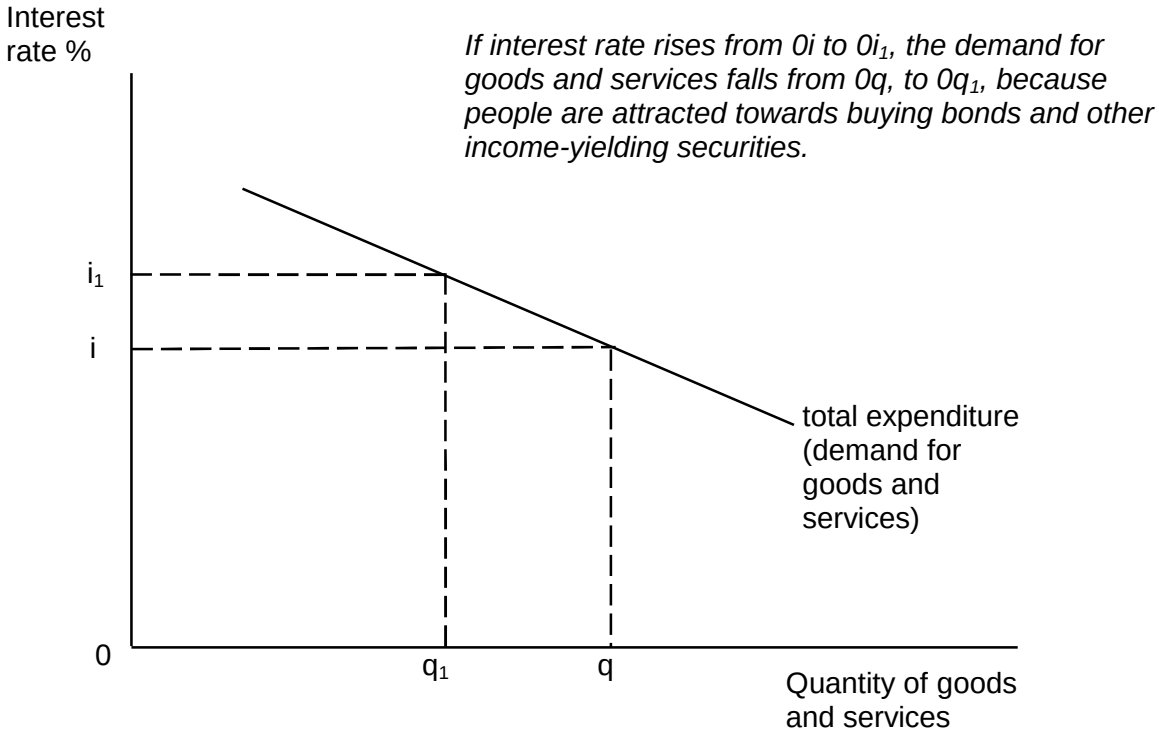


Figure 13.3: Monetarist view of demand and changes in interest rate

Classical and Monetarist View

We can now summarise the classical and monetarist position with regard to interest rates and money, and also with regard to interest rates and the demand for goods and services.

It is that the demand (and therefore the supply) of money is not very responsive to changes in interest rates. Putting this in more formal economic language: *money demand and supply are interest rate inelastic*.

On the other hand, the willingness to spend money on goods and services is responsive to changes in interest rates: the *expenditure demand for goods and services is interest rate elastic.*

The Keynesian View of Interest Rates and Expenditure

As we saw earlier, in the Keynesian view of the national economy, consumption (i.e. total expenditure on goods and services) is mainly dependent on income levels. In other words, the main influence on the level of consumer demand is seen as the level of income and not the supply or the price of money (interest rates).

Therefore the Keynesian does not believe that changes in interest rates are likely to have much effect on the level of expenditure (consumer demand). Again, the more formal economic statement is that *total expenditure or demand for goods and services is believed to be interest-rate inelastic.* In contrast, we have seen in this study unit that the Keynesian, stressing the speculative motive in liquidity preference, believes the *demand (and hence the supply) of money is interest rate elastic*.

Implications of the Differences

These two differing views of the relationship between interest rates, demand for money and demand for goods and services have major implications for government policy, especially for policy on money supply and the control of money supply.

© ABE and RRC

Suppose it is possible for the government to engineer a reduction in the money supply – e.g. by forcing the banks to reduce lending to customers – and so reduce their credit creation. Then this change in supply, like any other market shift, will result in a price change. Interest is the "price of money", so a reduction in money supply can be expected to force up interest rates. But the amount of change will depend on the supply and demand elasticities – on the responsiveness of supply and demand to interest rates. Given that there will be some effect on interest rates, this in turn will affect total demand for goods and services – again, the extent of effect will depend on the relationship between expenditure demand and interest rates.

Now we can begin to see the importance of the differences in views between Keynesians and monetarists. These are illustrated in Figure 13.4.

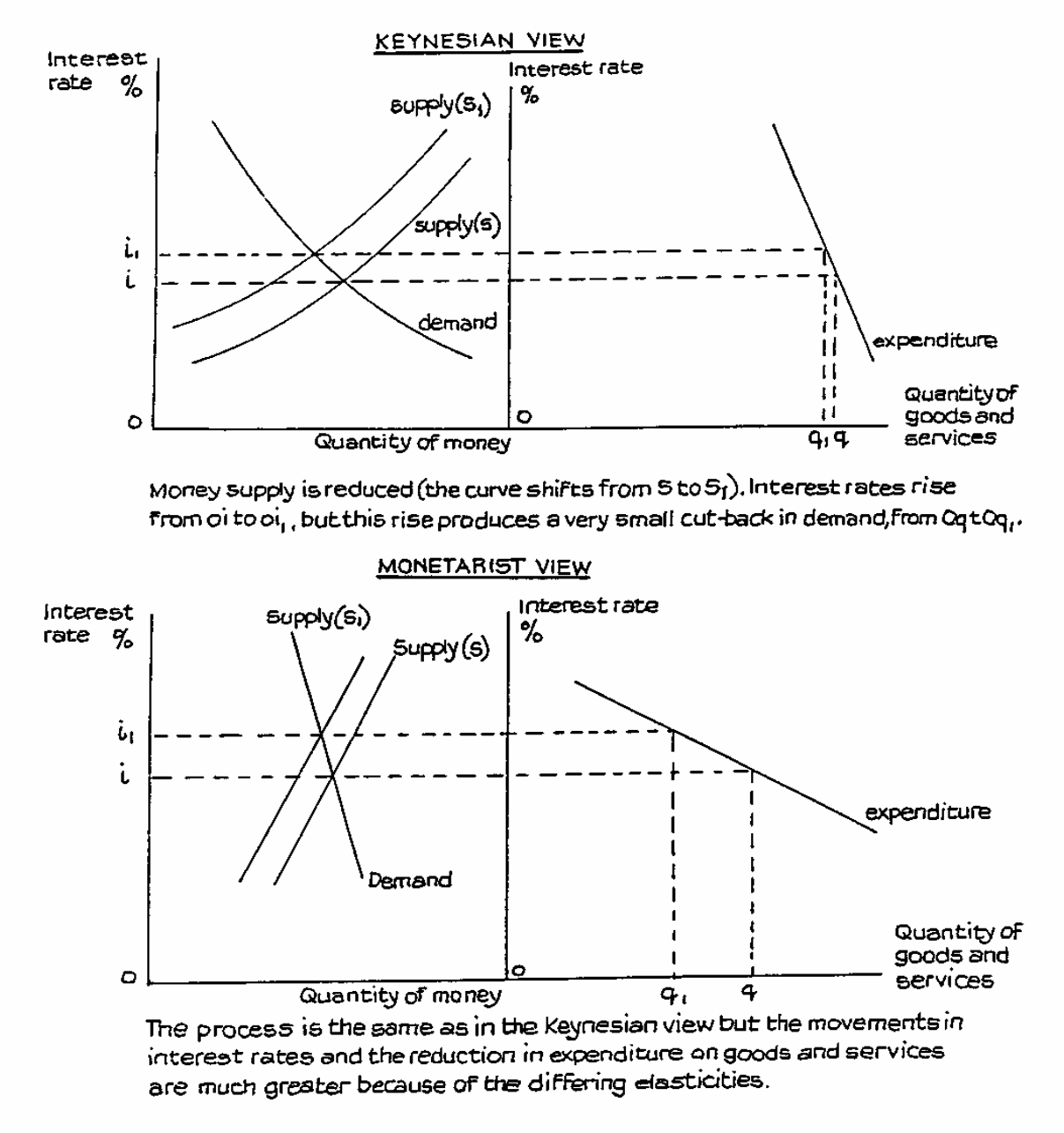


Figure 13.4: Keynesian and monetarist views

Keynesians believe that there is a close relationship between money demand and interest rates, but this interest rate elasticity ensures that any shift in rates brought about by a forced shift in supply also reduces demand: so in effect, the interest rate change is small. Expenditure is not much influenced by interest rate anyway (it being influenced more by income), and the small rise in interest produces little movement in expenditure.

The position according to the monetarist view is very different, although the mechanism is the same. Demand remains largely unaffected by the shift in supply and the change in interest rate, which is thus pushed up higher than in the Keynesian view. This steep rise in rate produces a major reduction in the interest-responsive demand for goods and services.

In effect these are very marked contrasts, and you would expect the debate to be settled fairly easily by research into actual interest rate and money supply changes. In practice, economists' research faces a great many practical difficulties. As we shall see, not least the problem of actually defining and measuring money supply and innovations that affect the demand for money in the financial system.

However, the Keynesian-monetarist controversy of the 1970s and 1980s is now more of interest to students of the history of economic thought, than for the understanding of monetary policy. The overwhelming weight of empirical evidence and practical experience in the conduct of monetary policy since the 1970s is that money matters, and monetary policy is effective as a means of controlling the level of aggregate demand and hence the rate of inflation.

D. CHANGES IN LIQUIDITY PREFERENCE

So far we have looked at the consequences of changes in the quantity of money demanded in response to changes in interest rates. We also need to consider the effect of a shift in the demand for money or the whole liquidity preference curve, i.e. see the effects when people wish to hold more (or less) liquid money at all relevant rates of interest.

If people desire to hold a higher proportion of their wealth in the form of liquid money, then they will have less available for use as loanable funds or to purchase physical assets. The logical consequences of reductions in each of these would be to reduce levels of business investment.

- If the supply of loanable funds falls, we would expect interest rates to rise. This would increase the investment costs faced by business firms and tend to reduce their investment intentions.
- If expenditure on goods and services falls, this would reduce the aggregate level of consumer expenditure and lead to a reduction in business investment. Firms invest in order to increase future production. There is no point increasing future production if current expenditure on goods and services is falling. The reduction in investment would have a depressing effect on the equilibrium level of national income through the investment accelerator and multiplier.

This process and the terms "investment multiplier" and "investment accelerator" are explained in Study Unit 10. At this stage it is simply necessary to recognise that any reduction in investment is likely to depress the general level of economic activity in a country.

© ABE and RRC

E. THE QUANTITY THEORY OF MONEY AND THE IMPORTANCE OF MONEY SUPPLY

The Money Equation

Changes in the supply of money in an economy can affect the rate of interest and hence the level of aggregate demand. Changes in the level of aggregate demand in relation to aggregate supply, as we saw in Study Unit 11, affect the general level of prices in the economy. Once the economy is operating at its full capacity/full employment level of output, additional injections of aggregate demand by means of increases in the supply of money will merely serve to drive up the level of prices. This provides the theoretical foundation for the central banks' use of monetary policy to control demand and the rate of inflation. This accords with the so-called monetarist view of money and inflation represented by *the quantity theory of money*. This, in very simple form, can be stated as follows.

$$MV = PT$$

where:

M = money supply or stock

V = velocity of circulation of money (i.e. speed at which it circulates between buyers and sellers)

P = average price of goods and services

T = number of transactions – i.e. volume of production
(T is sometimes written as Q, representing the quantity of production).

Now on its own, this equation tells us very little. However, the important issues lie in the relationships between the elements of the equation. Monetarists regard V as fixed or fairly fixed, and they also regard T (or Q) as fixed at a given level of technology. If these assumptions are correct, then effectively the two variables in the equation are M and P. A given change in M (the money supply) can be expected to produce a definite and predictable change in P (average prices). The relationship will not always be as simple as this, because allowance will have to be made for known variations in V and T, owing to forces outside the monetary relationship (e.g. improvements in technology and changes in the financial structure). It will also take time for any change in money supply to work through into general price increases, so that time lags of up to two years are suggested – though monetarists are not always in agreement over the precise time lag.

There is a further modification that many modern monetarists would make to this argument. This recognises that prices tend to be flexible upwards but not downwards: thus it is argued, if money supply is increased, then average prices will rise as already indicated; however, if money supply is reduced sharply, then prices do not fall. The variable that has to give in this situation is T (or Q), i.e. total output in the economy, as firms cut back production and consequently employ fewer workers. The implication of this is that an attempt to cure inflation by a sudden and sharp reduction in money supply will lead instead to an increase in unemployment rather than a check or reversal in price rises. The reasons for this "ratchet effect" for prices are that large firms are reluctant to reduce their product prices, and trade unions and workers resist strongly any suggestion of a reduction in wages.

Diagrammatic Representation of the Quantity Theory of Money

We can illustrate the monetarist analysis of the relationship between changes in demand and price quite simply, and this will also help to emphasise some of the assumptions on which the view is based.

© ABE and RRC

We must first repeat the belief that changes in demand arise from changes in money supply and the price of money. Remember that, all other things remaining equal, an increase in money supply can produce a reduction in interest rates, which in turn can lead to a significant increase in aggregate demand.

Look now at Figure 13.5, which illustrates the effect of an increase in aggregate demand. The economy is initially in full employment equilibrium (Ye), determined by the point of intersection of AD_1 with the LRAS curve at point E_1. The aggregate demand curve AD_1 is drawn on the assumption that the economy has a fixed supply of money MS_1. This assumption corresponds to that of a fixed M in the quantity theory equation. Yet the full employment level of output corresponds to the fixed level of total transactions T or production Q in the quantity theory equation. The position of the economy's LRAS curve can change over time with economic growth. However in the absence of economic growth, the economy's maximum level of sustainable real output or national income is fixed, and cannot be increased by increasing the level of prices in the economy. This is what is shown by the vertical LRAS curve, and is the same as the assumption made regarding the fixity of T or Q in the quantity theory of money.

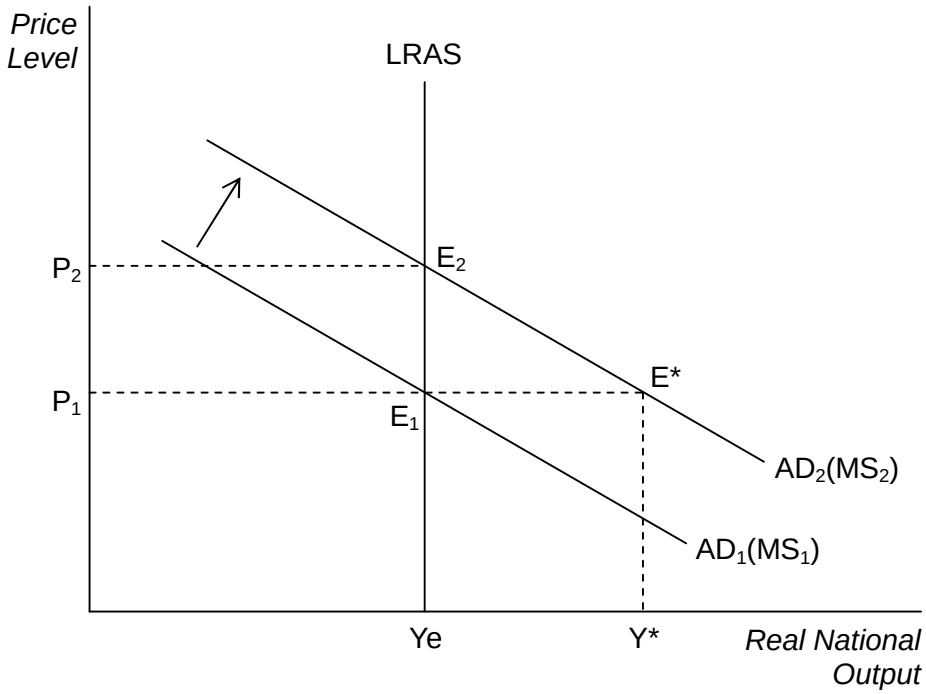


Figure 13.5: Increase in aggregate demand

Now assume that the central bank increases the supply of money in the economy from MS_1 to MS_2. All other things remaining unchanged, the increased supply of money will cause a reduction in the level of interest rates in the economy, as shown in Figure 13.2. The reduction in the level of interest rates will in turn lead to an increase in expenditure in the economy, as shown in Figure 13.3. The increase in expenditure is the same as an increase in the level of aggregate demand, and this is represented in Figure 13.5 by the shift to the right in the aggregate demand (AD) curve from AD_1 to AD_2. To indicate that the shift in the AD curve is the result of an increase in the supply of money in the economy, the two AD curves have their associated supply of money indicated by MS_1 and MS_2.

At the initial equilibrium price level P_1, following the increase in the supply of money the new level of aggregate demand in the economy is E* on AD_2. The level of aggregate demand at E* is Y* and this is excessive relative to the economy's capacity output Ye. That is, it lies to the right of the LRAS curve. Although the economy may be able to produce a higher level of

© ABE and RRC

output than Ye in the short run by operating on its initial SRAS curves (not shown in Figure 13.5 for clarity of exposition), the excess of aggregate demand in the economy will drive up the level of prices. Indeed, the economy will continue to experience rising prices (inflation in other words), until it reaches a new point of stable equilibrium at E_2 on its LRAS curve. The new point of equilibrium at E_2 corresponds to the prediction of the quantity theory of money. If the economy is subject to an increase in the nominal supply of money when it is already operating at full capacity, all that will happen once the extra demand created has worked its way through the economy will be a rise in the general level of prices in proportion to the increase in the supply of money. That is, in figure 13.5 the increase in the supply of money from MS_1 to MS_2 merely moves the economy up the LRAS curve from E_1 to E_2. The only change is an increase in the level of prices from P_1 to P_2.

F. METHODS OF CONTROLLING THE SUPPLY OF MONEY

Whatever the argument on the precise timing and severity of policies needed to control inflation, all monetarists believe that there has to be strong government control over the supply of money. In fact, even Keynesians would accept that there has to be some degree of control over money supply, though they would not elevate these controls to the important place claimed by monetarists. We must now look at some of the methods by which governments attempt to control the money supply. Remember that all our definitions of money have been based on deposits held by banks or similar financial institutions, so that you must expect control over money to appear as a form of control over the power of the banking system to create credit.

Interest Rate Control

Remember that money supply and demand are very closely related. If the price of money rises – i.e. if interest rates rise generally – then the demand for money can be expected to fall, although an interval may be necessary for the full effects to be felt. If people wish to borrow less, then the banks may be expected to lend less. If the banks lend less, then the volume of deposits will rise more slowly, and money expansion may be checked. A government or the central bank may therefore seek to control the supply of money through its ability to influence the level of interest rates. This is the main method of controlling the supply of money used by central banks in advanced economies.

Control over Banking Ratios

In Study Unit 12 we introduced the bank credit multiplier and saw how the proportion of customer deposits held as cash affects the lending power of the banks. If the proportion is one-tenth, then the multiplier is ten. If the proportion rises to one-eighth, then the multiplier falls to eight, and so on. A central bank may seek to influence the value of the bank credit multiplier by changing the value of the commercial banks cash reserve ratio. For example, to reduce bank lending and hence the rate of growth of the money supply, the central bank could increase the minimum ratio of bank cash reserves to deposits.

Direct Controls over Banks

The government, acting through the central bank, may require the commercial banks to keep their customer lending within stated limits, or to discourage certain forms of lending, or forbid lending for stated purposes. In a market economy or a mixed economy containing a substantial free market element, such controls are unpopular and difficult to keep in force for very long. They may be regarded as the first step towards total control of the banking system or complete nationalisation of all banks.

These methods of control assume that the central bank does not itself operate directly in the ordinary commercial finance markets. In some countries, the national central bank does lend

© ABE and RRC

directly to industrial and commercial organisations. In such countries, a government wishing to control the money supply would have to keep careful and strict control over these lending operations.

Control of Government Borrowing

A straightforward analysis of money supply and its changes suggests that an increase in government borrowing will increase money supply *only* if this is financed through the banking system. If it is financed by direct borrowing from the public, through sales of bonds, then there is no increase in money supply, and there could be a reduction through the withdrawal of money from private sector deposits with the banks to pay for the government securities. Thus there is a connection between fiscal policy and monetary policy. The effective control of the money supply and inflation requires governments to exercise fiscal discipline and limit their expenditure to what they can pay for out of tax revenue and borrowing from the public, not the banking system.

However, even in this case, there may be indirect consequences. If the government enters the finance market to compete for a larger share of private savings, then firms may be forced to borrow from the banks instead of raising money through issues of shares or debentures (long-term securities). This suggests that the government is *crowding out* private investment and forcing it into the banking system. Also, if the government forces up interest rates because it is competing with building societies and banks and capital markets for private savings, firms will be unwilling to incur long-term debt at high rates of interest. Instead they will prefer to borrow on short-term and on more flexible terms from banks, in the hope that future conditions will be more favourable for longer-term funding.

G. MONETARY POLICY AND THE CONTROL OF INFLATION

Money is important because a modern economy cannot function without an adequate supply of a sound medium of exchange and an efficient financial system. However, as the quantity theory of money demonstrates, an economy can have too much of a good thing in that excessive growth of the money supply merely leads to inflation. Changes in the supply of money can affect interest rates and the level of aggregate demand in the economy. If the economy is suffering from deficient aggregate demand, an expansionary monetary policy will lead to increased employment and real output. Monetary policy can be used in the same way as fiscal policy to regulate the level of aggregate demand. What monetary policy cannot do is create jobs and prosperity out of nothing. In a modern economy money is, after all, nothing but pieces of paper and electronic records in bank computers. Once an economy is operating at full capacity, its real output and citizens' standard of living is determined by its stock of physical and human capital, not its supply of money. Continued expansion of the supply of money in an economy thus eventually leads to inflation, not growth and prosperity.

One of the myths of economic development and growth is that they are both helped by inflation and impossible without it. In fact, the clear message of the study of economic growth in different countries is that there is no clear positive relationship between the rate of inflation and the rate of economic growth. Some countries have experienced high rates of growth and inflation, while other countries have suffered from high rates of inflation and economic stagnation. In contrast, some economies have enjoyed low inflation combined with high rates of real economic growth. What is established beyond any doubt is that once inflation becomes established in an economy it tends to accelerate and, if unchecked, eventually leads to economic disorder with falling output and increasing unemployment. High and accelerating rates of inflation affect economic behaviour and distort the effective working of markets. Because inflation erodes the value of money and undermines the logic of savings, it stimulates current consumption and speculative investment in physical assets, especially

© ABE and RRC

land and property. Avoiding loss due to inflation takes priority over creating new jobs and real wealth through productive investment in business.

The dangerous internal dynamics of inflation are due to the role of expectations. Once inflation becomes established, people try to avoid its costs by anticipating the future rate of inflation and taking appropriate avoiding actions. If the rate of inflation is expected to increase, the sensible thing to do is spend more and save less before the expected increase in the rate of inflation reduces the value of money and savings even further. But this merely increases aggregate demand relative to aggregate supply, and puts even more upward pressure on prices. This leads to the interesting conclusion that if people expect inflation to increase and act accordingly, the actual rate of inflation will increase as expected. This leads to self-reinforcing behaviour, or *self-fulfilling expectations*. Correctly anticipating an increase in the rate of inflation leads people to anticipate yet further increases, and this behaviour continues to fuel the acceleration in the rate of inflation. Once the rate of inflation starts to accelerate, and people expect it to continue accelerating, it becomes difficult to reverse people's expectations of its continuation.

Modern monetary policy is based on the view that inflation yields no permanent benefit for an economy and can cause much economic harm if unchecked. Once inflation is fully accepted in an economy, monetary policy looses all its power to do good but retains its power to cause yet more inflation. For this reason it is better to avoid high rates of inflation, and the problem of trying to reverse people's expectations of ever increasing inflation, by maintaining a very low rate of inflation and creating the expectation that the rate of inflation will stay low.

Monetary policy can be used to achieve monetary stability if the government or the central bank announces *a target for the annual rate of inflation,* and achieves the target by managing the level of demand in the economy through its control of the rate of interest. Countries that operate monetary policy on the basis of a target for the rate of inflation usually also have an *independent central bank*. Central bank independence refers to the removal of political control and interference from the conduct of monetary policy by the central bank. A *fully independent central bank*, such as the European Central Bank (ECB), sets its own target for the rate of inflation as well as operating monetary policy free of government influence in such a way as to achieve its target. It is of the upmost importance for the success of inflation targeting that the central bank is completely free of any control or influence from the government, because such interference would undermine people's confidence in the ability of the central bank to keep inflation under control at the target level.

For example in the UK, the government has set the target for the rate of inflation at two per cent, plus or minus one per cent. The government has given the Bank of England the task of achieving the target for the rate of inflation. To make sure that people believe that the Bank of England will achieve the target and keep the UK's rate of inflation close to two per cent, the government gave the Bank of England operational independence in 1997. What this means is that the Bank of England now operates as an *independent central bank*. The Bank of England is not fully independent, because the UK government still determines the target for the rate of inflation. But given the target set by the government, the Bank has complete autonomy. It is allowed to independently set a monetary policy to enable the economy to achieve the target rate of inflation. This means that the Bank of England sets the level of the rate of interest each month purely on the basis of the level required to control inflation and, more significantly, people's expectations of the rate of inflation. An independent central bank sets interest rates at the level required to achieve the target rate of inflation even when the government, for either valid or politically motivated reasons, would prefer the central bank to set the rate of interest at a different level. If the central bank's independence to determine monetary policy is compromised by political interference, then public confidence in the achievement of a low and stable rate of inflation is likely to be destroyed. Once the belief in an effective anti-inflation policy is lost, the public will start to anticipate accelerating inflation and inflation will return to undermine employment, output and living standards.

© ABE and RRC

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. Explain the meaning of the demand for money (liquidity preference).
2. Explain, using a demand for money curve diagram, why the demand for holding money decreases as the rate of interest increases.
3. Outline the quantity theory of money
4. Explain how a central bank controls the level of short-term interest rates in the economy.
5. How is the effectiveness of monetary policy affected by:
   (i) the interest sensitivity of the demand for money, and
   (ii) the interest sensitivity of investment expenditure?
6. What is an "independent central bank"?
7. What is an "inflation target"?

© ABE and RRC

© ABE and RRC

Study Unit 14

Macroeconomic Policy

Contents ***Page***

A. The Major Economic Problems 250
What is an Economic Problem? 250
Inflation 250
Unemployment 250
Trade Difficulties 251
Regional Problems 252
Lack of Adequate Economic Growth 252

B. Policy Instruments Available to Governments 253
Fiscal Policies 253
Demand Management and the Deflationary Gap 255
Demand Management and the Inflationary Gap 256
Monetary Policies 256
Direct Controls 256
Government Spending 257

C. Policy Conflicts and Priorities 258
Difficulties in Pursuing all Objectives at Once 258
Differences in Priorities 258

D. Supply-side Policies 259
The Natural Rate of Unemployment 259
Supply-side Objectives 260
Taxation and Fiscal Measures 261
Trade Unions and Supply 262
Encouragement of Competition 263
The Removal of Bureaucratic Controls over Business 263

© ABE and RRC

Objectives

The aim of this unit is to explain and interpret the main objectives of government macroeconomic policy.

When you have completed this study unit you will be able to:

- explain the conflicts that can arise between various macroeconomic objectives
- use aggregate demand and supply diagrams to demonstrate how these conflicts arise
- discuss the possible advantages of using fiscal and monetary policy together to try to reconcile conflicts in macroeconomic objectives
- show, using aggregate demand and aggregate supply diagrams, how the degree of underutilisation of an economy's labour resources can affect the trade-offs between economic growth, price stability, interest rates and unemployment.

A. THE MAJOR ECONOMIC PROBLEMS

What is an Economic Problem?

In many respects an economic problem, as perceived by a government, is an aspect of what is generally known as *the fundamental economic problem*:

> the attempt to satisfy unlimited wants with scarce resources, so that full satisfaction is impossible and choices have to be made between competing claims on those resources.

At the same time, this general problem is aggravated by inefficiencies in the production system, so that the achievement of available resources is not as great as it might be.

In practice we can identify a number of distinct problems which afflict modern industrial economies, and which are considered to be within the power of modern governments to reduce, if not totally solve.

Inflation

As we have already noted, inflation is the term used to describe a condition of constantly rising product and factor prices – the main factor price being wages: the price of labour. Inflation is a problem because it makes the production and distribution system less efficient. It creates uncertainties about costs, and it makes planning more difficult and uncertain. It makes long-term agreements difficult to make, because past agreements become unjust as the value of any agreed constant payment is steadily reduced. Money is unable to fulfil those functions which depend on confidence that it will retain its purchasing power and acceptability in the future. Savings lose their value, and people who have saved for future needs feel a sense of injustice. Countries suffering the most severe rates of inflation find that their exports become more expensive and difficult to sell in world markets, while imports become cheaper and grow in volume.

If inflation is not checked, it increases in intensity until prices rise daily and all confidence in money is lost. Trade reverts to a basis of barter, and all confidence in the financial system collapses. This condition of *hyperinflation* is usually associated with extreme political and social unrest and uncertainty for the future.

Unemployment

Unemployment is said to exist when *resources*, especially people available and seeking work, cannot find employment. It is an economic problem, in the sense that the community loses the production that could have been achieved, had all resources been employed.

© ABE and RRC

Unemployment is also a major social problem, because work is an important element in a person's standing in the community. A person who feels that he or she ought to be working but who cannot find work often feels rejected by society and, not uncommonly, resorts to antisocial behaviour.

We have already noted that Keynesians and monetarists have differing views concerning the nature and causes of unemployment, and it is convenient here to summarise some of these important differences.

- Both groups agree that there are elements of frictional and structural unemployment, but monetarists believe that the structural element may be artificially higher than it need be. This is because high state unemployment and welfare benefit payments reduce the pressures to adjust to changing economic conditions. They also believe that social attitudes and selfish protectionist motives by trade unions delay adjustment to change.
- Keynesians believe that much unemployment is caused by a deficiency in total demand consisting of household spending (C), business investment (I) and government spending (G). This is termed "demand deficient unemployment". Monetarists also believe in the possibility of demand deficient unemployment, but that both monetary and fiscal policy are less powerful than argued by Keynesians in combating such unemployment. Monetarists believe that the more effective government policies are ones aimed at preventing the emergence of demand deficient unemployment. Keynesians, on the other hand, focus on policies for reducing demand deficient unemployment once it has happened, rather than looking to policies to prevent its emergence.
- Monetarists believe that the natural rate of unemployment would be very small, if markets were free to operate according to the unrestricted interplay of the normal market forces of supply and demand. The effective functioning of markets requires a low and stable rate of inflation. For this reason, monetarists see the effective control of inflation as an essential precondition for preventing the emergence of demand deficient unemployment. The natural rate of unemployment is that rate which exists when the total demand for labour is roughly equal to total supply. People are then unemployed for frictional reasons – the normal wear and tear of firms closing, people changing jobs for personal reasons and so on, and for structural reasons – changes in the labour market caused by shifts in product demand and changes in production technology. The supply side of the economy is as important as the demand side in preventing a high rate of unemployment. This means that governments also need to use supply-side policies and encourage investment, education, labour flexibility, mobility and skills training to boost productivity growth. A high rate of unemployment is therefore blamed on imperfections in labour markets, barriers to productivity growth and inadequate investment in physical and human capital. These are seen mainly in terms of failure to understand and to adjust to structural change, undue trade union power and the system of government taxation and benefits payments that penalises work effort and entrepreneurship. Monetarists argue that a large part of high unemployment is voluntary. This is in the sense that people are waiting for jobs they think suitable instead of accepting what is available, and because they support trade union measures which force wages above the market equilibrium and so reduce the demand for labour.

Trade Difficulties

Trade difficulties are closely associated with inflation which increases export and reduces import prices in world markets. Both Keynesians and monetarists would agree that rising imports indicate a condition where demand is greater than the supply from the home production system. However, whereas Keynesians would concentrate attention on what is perceived as excess demand, monetarists pay more attention to failings in the supply or

© ABE and RRC

production system. Monetarists would tend to regard this as inefficient for a variety of reasons, including trade union power, lack of profit incentives, inefficient management and often being associated with monopoly power and bureaucratic barriers to business enterprise.

Trade difficulties are manifest in the structure of a country's balance of payments accounts and are usually associated with deficits on the current account of the balance of payments. However, in many cases a country's balance of payments problems are compounded by the choice of inappropriate exchange rate policy.

Regional Problems

If you live and work in the United Kingdom, you will probably be aware that the central problems of inflation and unemployment do not affect all areas of the country with equal intensity. In the southern areas inflationary pressures seem to be greater, whereas unemployment is generally more severe in the northern areas. If you lived in some other country, you would probably be aware of similar regional differences. These are regarded as economic problems, because the failure of some areas to develop as successfully as others suggests that production is being lost through the underuse or inefficient use of available scarce resources.

People tend to think that they are well or badly off, according to the comparisons they are able to make with other people. If living standards and employment opportunities are very different in different regions, there is likely to be social and political discontent. There is also the problem that large-scale movement of people from one region to another to find employment is a further possible cause of social unrest. Families are divided and pressures build up on housing and other services in the more prosperous areas.

If you do not live in the UK, you should try to apply similar general principles and arguments to the problems of your own country.

Lack of Adequate Economic Growth

What is adequate depends on what is achieved elsewhere. If the economy of the UK grows at the rate of one per cent per year, this will be seen as inadequate if other countries of similar size and stages of development are able to achieve growth rates of four per cent or more.

It is also true that all the problems identified in this study unit so far seem fairly minor if the economy is growing at what is seen as a fast rate, and if living standards for the great majority of the people are rising fast and constantly. On the other hand, if there is very little growth, then these problems become magnified and harder to solve. People's aspirations may be raised by what they see being achieved in more successful economies, and there is dissatisfaction and unrest at the failure to make similar progress at home. When there is a high rate of growth, governments have resources to introduce measures which are politically popular, and their chances of keeping power are greater. Low growth and inability to carry out popular measures make it difficult for governments to stay in power, at least by democratic means.

© ABE and RRC

B. POLICY INSTRUMENTS AVAILABLE TO GOVERNMENTS

Fiscal Policies

Fiscal policies relate to the use of government spending and taxation as instruments to influence the economy. The original idea behind fiscal policy was chiefly associated with Keynesian ideas of using the power of governments to influence aggregate demand. The assumption was that the economy is demand led, i.e. that total supply responds to changes in total demand. It is now recognised that the supply side of the economy is just as important as the demand side; the neglect of aggregate supply and inflation by Keynesians created an over-optimistic view of the power of government fiscal policy. The Keynesian 45 degree model of income determination overstates the effectiveness of fiscal policy, and for this reason the more realistic *aggregate demand and supply model* is used.

Figure 14.1 illustrates an economy suffering from demand deficient unemployment. The economy is in equilibrium at E_1 with a level of real national output of OY_1. The consequent differences between what could be produced at the level of OYe, the full employment output level, and the actual level of OY_1, creates a deflationary gap represented by a – b. As long as this gap remains, there will be unemployment, caused by the deficiency of aggregate demand.

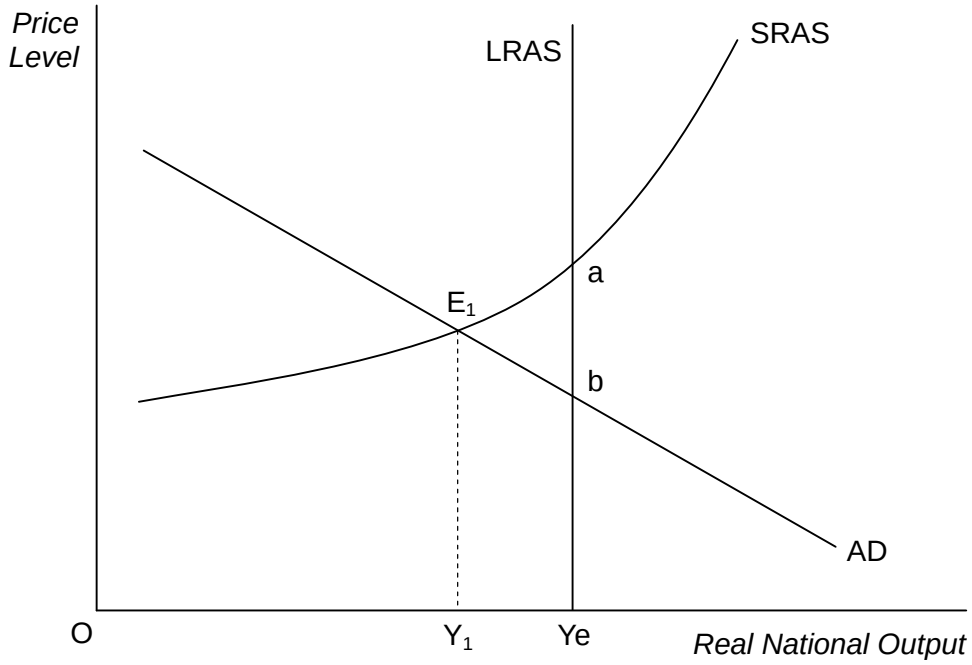


Figure 14.1: Economy suffering from demand deficient unemployment

Clearly the remedy this analysis suggests is to raise the aggregate demand curve by an amount equal to a-b in order to remove the deflationary gap. This is shown in Figure 14.2.

© ABE and RRC

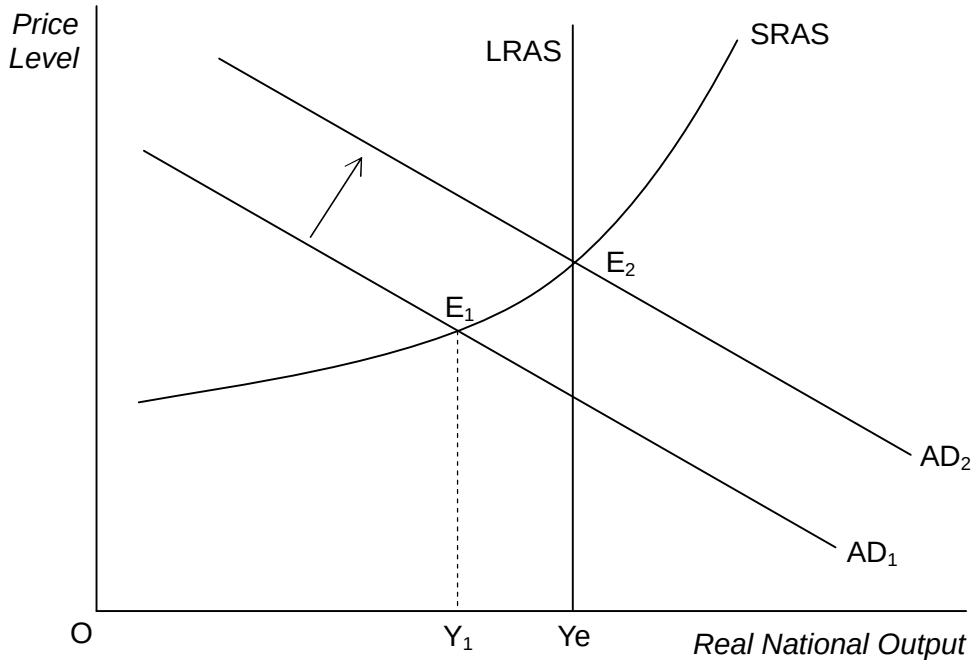


Figure 14.2: Removing the deflationary gap

The increase in aggregate demand in the economy is shown in Figure 14.2 by the movement of the aggregate demand curve from AD_1 to AD_2. This increase in aggregate demand moves the economy to a new equilibrium at point E2, where the SRAS curve intersects the LRAS curve. The new equilibrium is on the LRAS curve, which means that the deflationary gap has been eliminated and unemployment in the economy has been reduced. The movement of the aggregate demand curve from AD_1 to AD_2 could have been achieved by the use of fiscal policy or monetary policy. The diagram illustrates the consequence of the expansionary policy, not the policy measures used by the government to achieve the increase in aggregate demand. This means that the same diagram can be used to analyse the working of both fiscal and monetary policy.

In the case of fiscal policy the increase in aggregate demand, represented by the rightward shift in the AD curve in Figure 14.2, could have been achieved by the government increasing its own expenditure without increasing taxation to pay for the increase, or by maintaining its expenditure and reducing the amount of tax it collected. Either measure involves a deterioration in the government's budget position and an expansionary fiscal policy. Remember that in practice the government does not have to directly increase aggregate demand by the full amount of the initial deficiency. This is because the income multiplier will come into play, and the final increase in aggregate demand will be greater than the magnitude of the government's initial fiscal injection.

If the government uses monetary policy to boost the economy and eliminate the deflationary gap the diagram will look the same, but the rightward shift of the aggregate demand curve from AD_1 to AD_2 will result from an expansionary monetary policy. The central bank will undertake an open-market operation to reduce interest rates and increase the cash reserves in the banking system. The reduction in the level of interest rates will lead to increased bank lending and an even greater increase in the supply of money. The extra bank lending will be used by business and household borrowers to increase investment and consumption expenditure in the economy. As with the expansionary fiscal policy, the final increase in consumption and investment expenditure will be greater than the initial stimulus to demand caused by the increase in the money supply, due to the operation of the income multiplier process.

© ABE and RRC

If an increase in demand is to reduce unemployment, then it must be assumed that total supply – and therefore the demand for labour and other production factors – will rise in response to the change in demand. The move up the short-run aggregate supply curve to the point of full employment equilibrium on the long-run aggregate supply curve at E_2 will thus be associated with a rise in the level of prices in the economy or inflation. How much prices rise as unemployment is reduced through the use of an expansionary fiscal policy clearly depends upon the slope of the short-run aggregate supply curve. It also depends upon the position of the short-run aggregate supply curve. This is because any increase in the level of money wages and other cost of production as demand expands will cause the short-run aggregate supply curve to shift upwards, and add further pressure to the rise in prices. The upward pressure on prices will also affect inflationary expectations which, if unchecked, may lead to further upward pressure on wages and other costs as workers and firms seek to protect their wages and profits from erosion in value due to inflation. Suddenly, the scope and ease of fiscal policy alone to restore the economy to full employment without creating a worsening situation of inflation looks less certain than suggested by the simple Keynesian model of income determination. Worse, the aggregate demand and supply curve model shown in Figures14.1 and 14.2 assumes that the government knows precisely the position of the LRAS curve and the exact extent of the deficiency of aggregate demand in the economy. If the unemployment is a result of a decline in the economy's productive potential, a leftward shift in the LRAS curve, fiscal expansion may create excess demand and an inflationary gap!

Demand Management and the Deflationary Gap

If our theory suggests that to reduce unemployment we must raise total demand, then the problem becomes one of how to achieve this. Our earlier national income analysis suggests that this can be achieved by injections of new demand, which will then be multiplied within the economy to produce the new and higher equilibrium level that is desired.

Keynesians argue that the desired effect can be achieved if the government is prepared to operate with an unbalanced budget, i.e. if it spends more than it receives in taxation. This means that to raise the total level of aggregate demand, the government can increase its own spending (G) without increasing taxes, and/or reduce taxes in order to encourage household spending. Remember that Keynesians believe that the most powerful influence on total spending is income. A reduction in income tax will increase people's net disposable income, so that they will increase their spending in accordance with the *marginal propensity to consume.*

Fiscal policies are thus very important to the Keynesian. It is through the adjustment of taxation, and income tax in particular, that the government is able to influence the level of disposable income. Such changes in taxation will have an immediate effect on spending and hence on aggregate demand, which in turn produces a change in supply and the level of employment. The effect of the income tax reduction does not end there. The initial injection of extra spending will produce a larger change, in accordance with the national income multiplier. There will also be an additional impact resulting from the perception by business firms that demand for their goods and services is increasing. To meet the increased demand, they will increase investment: this produces a further injection in the economy, with a further multiplying effect. The combination of investment accelerator and national income multiplier will ensure that the total increase in demand will be larger than the initial injection achieved by the tax reduction.

So the Keynesian relies on a fiscal policy of tax manipulation combined with a willingness to tolerate an unbalanced budget to achieve and maintain full employment – or something as close as possible to full employment – in the economy.

A reduction in indirect or expenditure taxes would also be expected to stimulate the economy, because more of the consumer's gross spending will actually go to the suppliers of

© ABE and RRC

goods and services. Firms can be expected to increase the quantities they are willing to supply at each level of market price – the supply curve will shift to the right. The precise effect on output and price will depend on the slopes of the demand and supply curves. This is analysed later in this study unit, but we can see that we would certainly expect some increase in supply and employment following a reduction in indirect taxes. Such a reduction could also mean the government having to be prepared to operate an unbalanced budget, with revenue falling short of expenditure and the difference made good by borrowing.

Demand Management and the Inflationary Gap

Theoretically there is no reason why fiscal policies employed to reduce a deflationary gap cannot be reversed to reduce an inflationary gap. This would mean reducing government spending and increasing taxes, using any excess of tax revenue over expenditure to reduce the national debt. However such policies meet serious constraints in practice. Attempts to cut public sector spending may provoke fierce political resistance and will be politically unpopular. We are all in favour of reduced government spending in general, but we all oppose any cuts in those areas of spending that affect us and from which we benefit! Similarly we all agree that taxation is necessary but we all dislike paying tax ourselves. Consequently it is much easier for governments to reduce taxation and increase spending than to raise taxation and reduce spending. It is therefore no surprise that Keynesian demand management policies have been more successful in reducing deflationary than inflationary gaps. As inflation came to be perceived as the major economic problem of the 1970s and 1980s, attention turned away from fiscal policy and towards monetary policy.

Monetary Policies

The theoretical basis of monetary policy, the money equation and the main elements of monetary controls were examined in Study Unit 12. You should make sure you understand how these differ from fiscal policies. Remember that monetarists and Keynesians share a common belief: that the major cause of inflation is an excess of demand over available supply. However the Keynesian belief that demand is mainly a function of the level of income has led traditional Keynesians to rely chiefly on fiscal measures, and led later Keynesians to support direct controls over the level of incomes. In contrast monetarists believe that demand is mainly a function of the availability of money and credit (money supply), and this has led to their reliance on monetary policy. Monetary policy involves the central bank's ability to control of the supply of money to determine the level of short-term interest rates in the economy and the publics' expectations regarding the future rate of inflation. A successful monetary policy will achieve a low and stable (and hence predictable) rate of inflation. This in turn will give the central bank the power to influence the level of long-term as well as short-term interest rates.

Direct Controls

A government can always obtain the legal powers to control certain aspects of the economy, but it must be remembered that these powers are usually only negative. A government can *prevent* people or firms from doing certain things, but it has considerable difficulty in forcing them into positive action – i.e. actually to do things it wants done – purely by the exercise of its legal powers.

Because of the failure of governments to recognise the limitations of Keynesian demand management fiscal policy, overambitious use of expansionary fiscal policies in the 1960s and 1970s led to ever higher rates of inflation and the associated problems created by accelerating inflation. In the UK the denial of the role of money supply in fuelling inflation compounded the problems created by excessive levels of government expenditure in relation to taxation and the economy's supply capacity. Faced with worsening inflation, the government resorted to the use of direct controls over prices and incomes. This was an attempt to control inflation without reducing government expenditure and the rate of creation

© ABE and RRC

of money used to finance government deficit expenditure. Not surprisingly, such direct controls failed because they confused cause and effect. A persistent acceleration in the rate of inflation is impossible without new money creation. Continued growth of government expenditure financed by expansion of the money supply will lead workers to demand higher wages and firms to keep on raising prices. This behaviour is not the cause of the inflation but simply a rational response to the inflation caused by the government's own policies. By the 1980s the truth dawned that direct controls were ineffective because they addressed the symptoms not the cause of the problem of wage and price inflation. Direct controls, especially over wages and prices, are now recognised as inappropriate and ineffective in a market economy, and are now no longer used by governments.

One of the most controversial examples of the use of direct controls by the British government has been the successive attempts made to regulate wages, and sometimes prices. Regulation of price or factor price without also controlling the forces of supply and demand is never successful, because it must lead to serious distortions in supply and demand and it threatens to destroy the whole mechanism of the market. During all the periods of attempted wage regulation, employers and unions found ways of overcoming the controls in order to keep the labour markets working. Even so, shortages of skilled workers sufficient to hold back the expansion of some profitable firms and industries have been blamed on these controls. Such shortages made it difficult for firms to attract workers into activities requiring long and difficult periods of training, when wages nearly as high could be obtained from less demanding work. Nevertheless, the pay of people employed in the public sector, which is largely insulated from the forces of supply and demand, continues to cause problems. There does appear to be a need for guidance from some kind of authority for public sector pay. As long as there are not generally agreed principles and the government simply relies on its power as an employer, continued disputes and feelings of injustice are highly likely.

Few economists believe that controls over wages and prices can ever be effective in a free market economy. Such controls usually attempt to deal with the symptoms, not the cause of the problem. Not only do they tend to cause new problems of their own, but the problems they are intended to deal with, especially inflation, are invariably caused by governments themselves. Governments generally think that they have more power than they actually possess. When controls are imposed to prevent actions that people would otherwise take, there will be attempts to evade the controls, and the government may be forced into increasingly difficult, complex and expensive control measures. For instance, many countries have sought to impose strict import controls, only to discover that they have created a major smuggling industry. Many of those responsible for maintaining the controls simply use their powers to increase their personal incomes with bribes from both legal and illegal traders. We have only to note the problems of seeking to prevent the import of illegal drugs to see what happens when a government tries to suppress trade for which there is an effective demand. It is only too clear that a government cannot stop the abuse of drugs just by trying to prevent drugs imports.

Government Spending

Government (public sector) spending is a major part of total demand, so that variations in government spending can be used to influence the level of national income and product. The Keynesian uses government spending as a "counter-cyclical" instrument. The government can inject additional demand when household consumption and business investment are considered to be too low, and reduce public sector spending when the economy is thought to be overheating with excess demand from the private sector. In practice, it is easier for governments to increase public sector spending than to reduce it.

The monetarist, while recognising the use of public sector spending as a means to regulate the total level of economic activity, wishes to keep the total of this spending as low as possible.

© ABE and RRC

However, both Keynesians and monetarists do agree that the pattern of economic activity can be influenced by government spending decisions. For instance governments have sought to encourage the development of the computer industry, by assisting investment and by helping schools to buy British-made computers. Government can influence the development of transport by spending on roads rather than on the railways. It can also try to help particular regions by directing some public activities to them, and away from London.

However governments are generally limited in what they can do. For example, a government may stop a firm from building a new factory in a particular place. However, if the firm says that if it cannot have the factory where it wanted it, then it will not build a new factory at all, there is very little the government can do. Similarly, a government may prohibit the import (and sometimes the export) of particular goods or goods from or to particular countries, but it cannot force people in foreign countries to buy goods made by its producers.

C. POLICY CONFLICTS AND PRIORITIES

Difficulties in Pursuing all Objectives at Once

Economists recognise the possible conflict of objectives in government macroeconomic policy. The success of demand management depends on holding a very fine balance between total demand and total supply, and any swing in one direction is going to lead to difficulties. In order to reduce unemployment, the Keynesian will wish to expand demand, and he would be prepared to operate an unbalanced budget. He accepts that this may bring about some price inflation, and that it could also lead to rising imports and trade difficulties. So to reduce unemployment, the Keynesian recognises that he may increase problems of inflation and excess imports. Similarly, he will accept that action to bring trade into balance, or to reduce the rate of inflation, will probably bring about a reduction in the growth of the economy and in an increase in unemployment.

A monetarist will have a rather different analysis. He believes that in the long term, successful achievement of economic growth, successful trade and full employment all depend on an absence of inflation and a stable financial system. He believes that business enterprise, freed to operate in unregulated markets, will achieve growth, exports and employment, provided that the government keeps its own spending under control, keeps a tight grip on the money supply, and avoids inflation. There is therefore no fundamental conflict of aims in the monetarist analysis in the long term. However, starting from a position of high inflation caused by misguided demand-management based on over-expansionary fiscal and monetary policy measures, the monetarist believes that it is not possible to avoid some increase in unemployment. The monetarist is also sceptical concerning Keynesian remedies for regional problems, as explained in the next section of this study unit.

The monetarist does not believe that macroeconomic policies, as understood by the Keynesian, are effective at all. The Keynesian is concerned with aggregates, in the belief that injections of demand from government spending and tax reductions will operate on the economy as a whole, to increase employment. The monetarist is not convinced that the government has the power to influence the whole economy in this way, and he tends to prefer supply-side policies which operate on the economy through improving the operation of individual product markets – i.e. through microeconomic measures. If all or the majority of individual markets operate more efficiently, then the economy as a whole will prosper.

Differences in Priorities

If to begin with, we adopt the Keynesian position, then it is clear that there has to be some sense of priorities in choosing objectives. This is because not all can be pursued at once. The Keynesian would argue that his most important objective is to achieve and maintain full employment – but that this may have to be modified from time to time if inflation or trade

© ABE and RRC

difficulties become too serious. However, the main objective is always to avoid large-scale unemployment and, if this is threatened, some inflation or trade imbalance may have to be accepted.

Critics of Keynesian economics would suggest that in practice, governments do little more than react to a series of crises. They lurch between expansion and deflation as each problem becomes steadily more serious, and as the production system becomes increasingly dislocated by sudden shifts in demand policy. They see the inevitable consequence as uncontrollable inflation, which eventually brings about mass unemployment as the production system fails to compete with more efficient foreign systems.

The monetarist thus argues that *there is no alternative* to controlling inflation and freeing private sector markets from controls and barriers, so that they can expand production and increase employment. In the meantime the effect of reducing public sector activity and restoring a more competitive and efficient private sector is likely to cause strains and to increase unemployment. We have seen earlier that monetarists differ in their approach to the timing of policies. Some prefer a gradual approach, accepting that inflation should not be brought down too swiftly, in order to avoid the social and political upsets of too rapid a rise in unemployment. Others consider that the adjustment can be carried out more quickly and that more vigorous methods can be applied to remove restrictions to industrial markets.

D. SUPPLY-SIDE POLICIES

The disappointing experience of demand management policies when inflation became a major economic issue, and the monetarist argument that demand expansion almost invariably led to inflation because of the failure of domestic production to respond quickly enough to demand stimulation, led to the development of what became known as supply-side economics. It is monetarists who are most closely associated with modern approaches to the stimulation of supply. In this approach, supply-side economics is seen as the use of microeconomic incentives to change the level of full employment, the level of potential output and the natural rate of unemployment. The objectives are to increase total production, to increase the productivity of labour, and to make producers more competitive in world markets. A government pursuing supply-side policies wants business firms to produce more and to employ more labour – but to do so profitably, in competitive markets.

The Natural Rate of Unemployment

Central to understanding the theory on which supply-side economics is based is the concept of the *natural rate of unemployment*. This is the rate at which the labour market is in equilibrium – i.e. in which labour demand is equal to labour supply, so that there are no pressures to increase or decrease money wages.

The natural rate of unemployment will never be zero, because at any given time there will be unemployment arising from two important causes. These are known as *frictional* and *structural* causes. Frictional unemployment arises from the normal wear and tear of business life. There will always be people changing jobs, for a whole range of different reasons. These will include dissatisfaction with an employer or with working conditions, moving home, the failure of individual firms or just simply boredom or the desire to do something different. It is not always possible to move immediately from one job to another, although the average length of time that a frictionally unemployed person can expect to be without work varies with the level of total unemployment. It is not usually more than a few weeks.

Structural unemployment has two related meanings. On the one hand it arises from shifting patterns of demand. For example if many women decide to give up wearing jeans and trousers, and instead choose to wear skirts and dresses, then jeans manufacturers will have to lay off workers, while skirt manufacturers will be expanding their activities. Different firms

© ABE and RRC

in different localities may be involved, and it is not always possible for workers in the declining activity to move quickly into one that is expanding.

The other form of structural change is also known as "technological change". It arises from changing production methods, usually from the increased use of machines, including advanced electronic devices and computer software which can do a great deal of work previously carried out manually. When this kind of change takes place, there is no immediate compensating expansion in another activity. New technology always creates new activities and occupations in time, but these may be very different from the old, requiring new and different skills, and they are often located in completely new areas. Structural unemployment from technological causes can be greater and more disruptive than that from shifts in demand.

The two types of structural unemployment are often related, in the sense that new technology creates new products which replace old ones. The transistor destroyed the radio valve industry; the small electronic calculator destroyed the production of slide rules and mechanical calculating machines. Modern electronics has changed a great deal of product demand, and it has had a very great impact on the labour market.

It is clear then, that if we regard the natural rate of unemployment as being made up of frictional and structural unemployment, it is likely to be much higher today than it was in the 1950s and the early 1960s, before the current electronics revolution. Where monetarists differ from other (and particularly Keynesian) economists, is in their belief that the whole – or almost the whole – of the actual amount of unemployment is natural unemployment. If the actual rate of unemployment is seen as being at a level which is socially and politically unacceptable – and economically damaging, in the sense that production that would be possible at a higher level of employment is being lost – then the problem lies in reducing this natural rate. Monetarists believe that this natural rate is too high, and that it can be reduced by microeconomic (supply-side) policies.

Clearly then, monetarists and supporters of supply-side theories take an almost opposite view to Keynesians on the basic causes of unemployment. Keynesians see unemployment and inflation as opposite forms of national income disequilibrium (the deflationary and inflationary gaps). Monetarist/supply-siders see unemployment and inflation as caused by similar forms of market failure, with inflation as the primary result of this failure, helping to produce unemployment by pricing domestic production and production workers out of employment in world markets. Much supply-side policy, therefore, depends on removing imperfections, including government intervention, from product and factor markets.

Supply-side Objectives

If you look at Figure 14.3 and bear in mind the earlier outline of objectives of supply-side economics, you will realise that supply-side policies will be designed to shift the supply of labour curve (SL curve) to the right – i.e. increase the number of workers prepared to work at each wage level. This will reduce the natural rate of unemployment, and move the actual demand for labour (and hence raise the production level) further to the right along the demand curve by reducing the gap between union-imposed and the market-equilibrium level of wages, and shift the demand-for-labour curve to the right by increasing employers' production intentions. A number of possible ways of achieving these results may now be examined.

© ABE and RRC

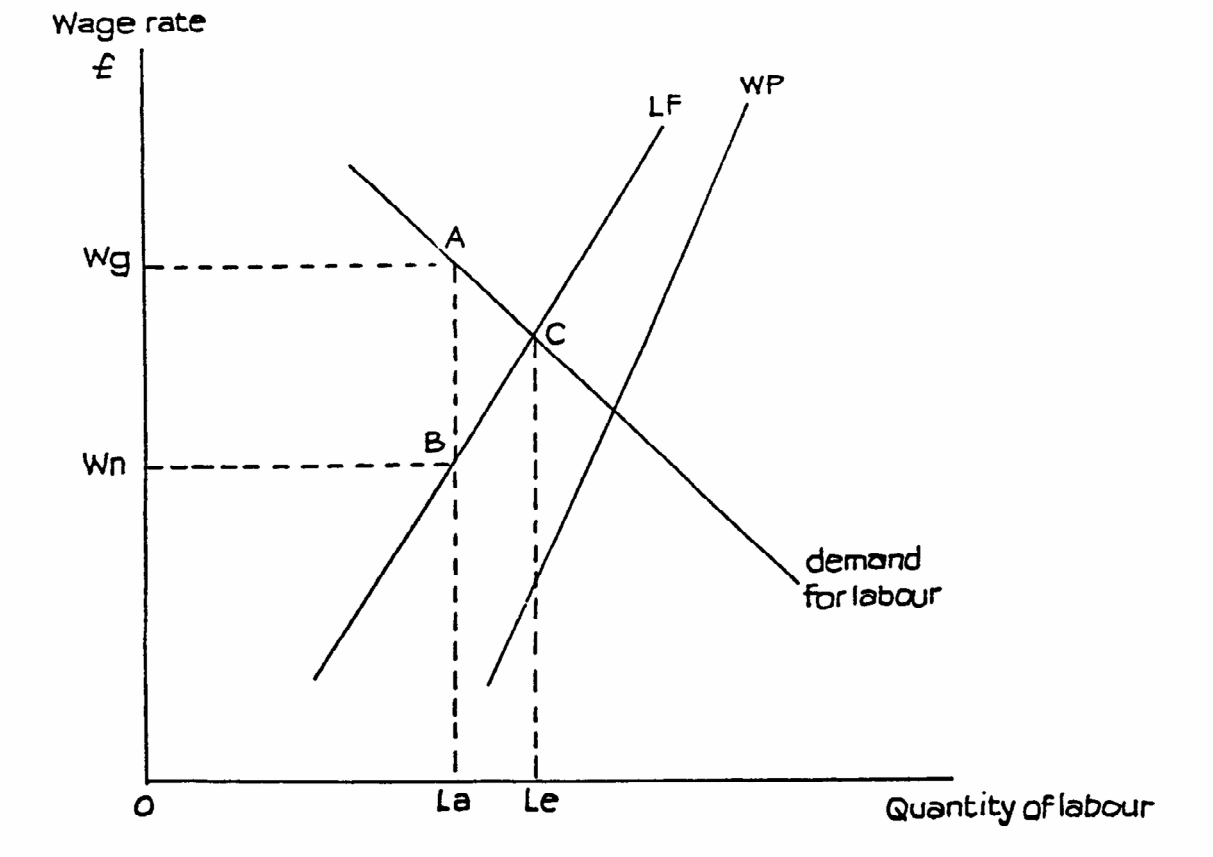


Figure 14.3: Effect of supply-side policies

Here are shown curves for the working population (WP), the actual supply of labour (LF) and the demand for labour, as before. If there are income taxes, and other payments of the nature of payroll taxes, then the wage cost may be OWg – the gross wage paid by employers plus compulsory payments which employers have to make, whereas the net wage actually received by the workers is OWn. The vertical distance AB represents the amount of income tax and payroll taxes. If this distance could be eliminated, the supply and demand for labour would move to the equilibrium position C, and employment would be at the higher level of OLe. Income and payroll tax reductions would have reduced the amount of unemployment by an extent depending on the slopes of the curves and the various distances involved.

Taxation and Fiscal Measures

As far as the public sector spending side of fiscal measures is concerned, there is a desire to reduce public sector spending in order to release resources of labour and capital for use in the private sector. This is because it is believed that private sector activity is more likely to generate further growth and employment, whereas much public sector activity and employment has to be paid for by taxation, which operates as a burden on the private sector and prevents its expansion.

The main objective is to reduce taxes both for employers and for employees. The effect of an income tax reduction for workers is illustrated in Figure 14.4.

In practice, the government recognises that this is impossible to achieve for the total labour market. However it may be possible for particular sections of the labour market which currently suffer from high rates of unemployment, especially in the markets for lower-paid and unskilled workers.

© ABE and RRC

If the pattern of income and payroll taxes is changed to reduce the burden on the low-paid workers, if necessary at the expense of the more-highly paid, the government will be able to avoid the criticism often levelled at tax reductions aimed at increasing labour supply. This criticism is that the supply-of-labour curve is backward-sloping, so that above a given wage rate, further increases in net wage will reduce rather than increase the willingness to work (because above a certain income level workers are more likely to prefer increased leisure to increased income). As long as the government's fiscal measures are concentrated on helping those whose net wage is below OW in Figure 14.4, which illustrates this concept, any achievement in increasing the net wage received by workers will raise the quantity of labour being offered to producers.

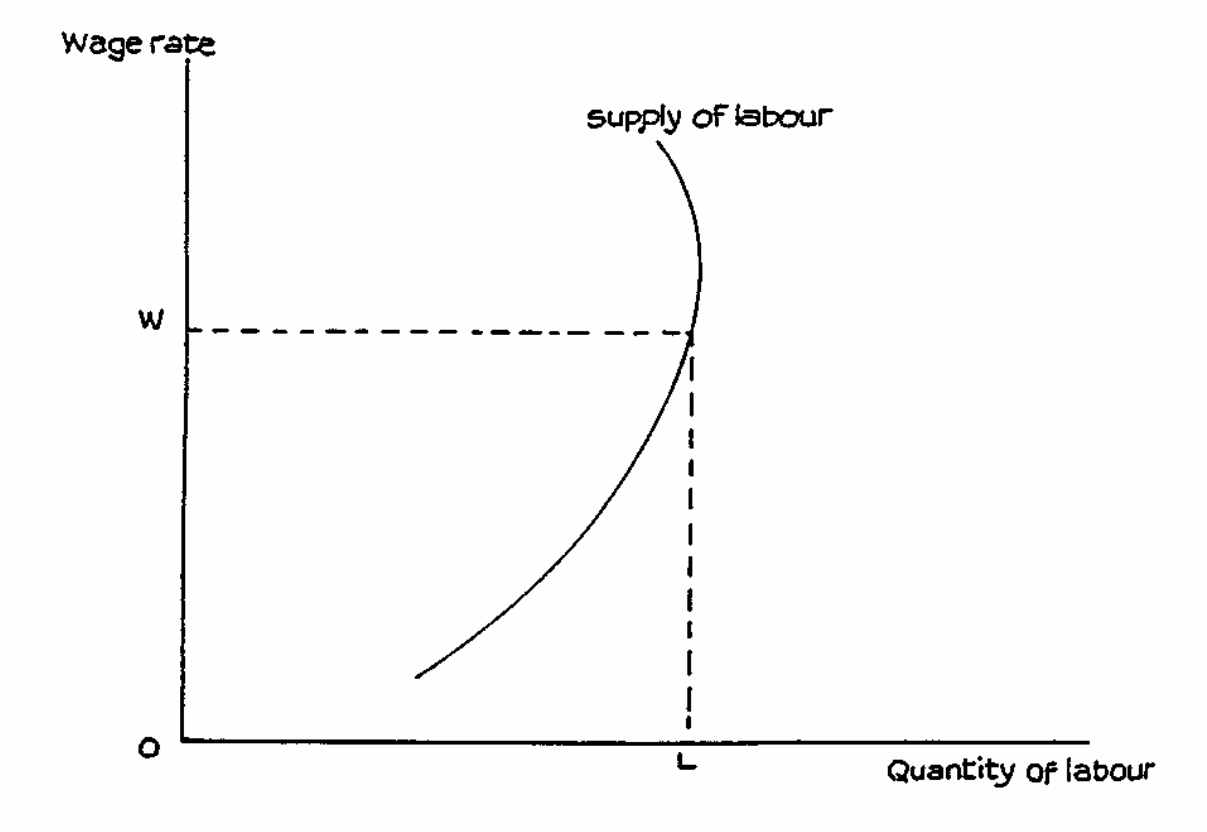


Figure 14.4: Effect of an income tax reduction for workers

Another aspect of supply-side fiscal policy is to increase the rewards of successful business enterprise. This is likely to involve a number of fiscal measures, including a reduction in the higher rates of income tax – i.e. the rates paid by high-income earners, on the assumption that a high proportion of these will be employers or business managers who are responsible for making the decisions that determine the level of output and for achieving business success.

Other aspects of tax reduction may involve granting tax allowances for investment in business enterprise by individuals, and reducing taxes on wealth and capital transfer. The latter would be regarded by supply-side economists as penalties imposed on people who have committed the "crime" of being successful in business, increasing the wealth of the community and the employment opportunities of others.

Trade Unions and Supply

To monetarists and those accepting supply-side theories, trade unions are generally regarded as being restrictive, reducing output, business profitability, competitive power, and increasing unemployment. Therefore any weakening in the power of trade unions might be

© ABE and RRC

expected to increase the ability of firms to survive and expand in competitive world markets, to make business production more profitable and therefore, desirable, and to reduce unemployment by allowing more workers to work at wages closer to the market equilibrium

Encouragement of Competition

Supply-side economists would regard the possession of undue market power by any organisation, whether worker or employer, as likely to reduce output and efficiency and raise costs and prices. Competition and the weakening of monopoly power is thus seen as a desirable objective, likely to lead to increased efficiency and production and, in the long run, to a higher and more secure level of employment.

The Removal of Bureaucratic Controls over Business

It is a frequent complaint of business managers in many countries, especially developing countries, that costs have been raised, efficiency reduced, and expansion hindered by the great range of planning and other bureaucratic controls to which business is subject.

Many controls on business activity are imposed for generally sound social reasons. These include the protection of the environment and the prevention of indiscriminate expansion of industrial activity, the protection of workers from exploitation, and the protection of consumers from unscrupulous or careless marketing and production. A defence can be made for such measures considered in isolation but, taken as a whole their cost can become too great. If the general result is to reduce output and employment, then the balance of cost and social benefit may have swung against the overall interests of the community. The problem is compounded further in two ways. In some cases the bureaucratic controls and restrictions benefit private interest groups in society at the expense of the rest, and the abolition of such restrictions is resisted strongly by those who benefit. The resistance to the removal of such barriers is often strong in developing countries, where the regulation and controls create widespread opportunity for bribes and corruption. The other reason why regulations and bureaucratic controls are damaging is that they can impede necessary change. The rate of technological progress increases dramatically and new products and productive processes require significant change in the structure of industry. The use of controls and restrictions to preserve old industries and ways of production, especially if faced with increased competition from imports because other countries have embraced the changes, may appear like a good way of preserving factories and jobs but it condemns the economy to longer term decline.

© ABE and RRC

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. List the four main macroeconomic policy problems a country may face.
2. Explain the following terms:
   - government budget deficit
   - fiscal policy
   - monetary policy
   - supply side policy.
3. Draw an aggregate demand and supply diagram to show how the government can use an expansionary fiscal policy to reduce demand deficient unemployment.
4. Draw an aggregate demand and supply diagram to show how the government can use an expansionary monetary policy to reduce demand deficient unemployment.
5. Draw an aggregate demand and supply diagram to show how the government can use a supply side policy to increase employment and reduce inflation.
6. Explain, with examples, why governments may face conflict between the achievement of the objectives of macroeconomic policy.

© ABE and RRC

Study Unit 15

The Economics of International Trade

Contents ***Page***

A. Gains from Trade and Comparative Cost Advantage 266
Common Advantages of Trade 266
Comparative Cost Advantage 267
Limitations to the Gains from Comparative Advantage 268

B. Trade and Multinational Enterprise 269
The Multinational Company 269
Reasons for Growth of Multinational Enterprise 269
Consequences of Multinational Enterprise 270

C. Free Trade and Protection 272
Advantages of Free Trade 272
Protection 272
Dangers of Trade Protection 275

D. Methods of Protection 276
Tariffs 276
Quotas 277
Embargoes 278
Voluntary Export Restraints 278
Export Subsidies and Bounties 278
Non-tariff Barriers 279
Exchange Control 279

E. International Agreements 279
Trading Blocs 279
GATT/WTO and the Liberalisation of Trade 282

© ABE and RRC

Objectives

The aim of this unit, in conjunction with Study Unit 16, is to explain the fundamental advantages and disadvantages of free trade, including the principles of absolute and comparative advantage.

When you have completed this study unit you will be able to:

- explain, using numerical examples, how gains from specialisation arise
- interpret data on opportunity cost
- identify economic reasons why governments may decide to promote free trade or impose restrictions on free trade
- explain the impact of free trade on business in developed and/or developing economies
- discuss the means that can be employed by governments to restrict or promote trade and evaluate the advantages and disadvantages of employing policies to restrict free trade.

A. GAINS FROM TRADE AND COMPARATIVE COST ADVANTAGE

Common Advantages of Trade

Even without any assistance from economic theory, it is not difficult to list some important advantages from international exchange. Among the more common benefits are the following.

(a) Better Supply of Goods

Through international trade, a country may obtain goods which it could not obtain otherwise. For instance Britain could not enjoy tropical fruit or manufactured goods made of copper, nickel, and many other metals, if it were not for the existence of international trade.

(b) Lower Costs

A country can obtain goods which it could not grow or produce itself, and it can also obtain goods which it could grow or produce – but only at higher cost than in other countries.

International trade, by opening up the whole world for trading purposes, increases the size of the markets for various goods. Production on a larger scale is then possible, allowing full advantage to be taken of economies of scale. For instance, if Switzerland only made watches for its own comparatively small domestic market, the cost of production per unit would be much higher than it is; in fact, Switzerland supplies many parts of the world with watches.

(c) Famines can be Prevented

World trade reduces the likelihood of famine and of other results of shortages of supply, since it is possible to offset temporary domestic shortages by getting additional supplies from abroad.

(d) A Curb on Monopoly

The existence of international trade is an obstacle to the development of monopolies. Even if there are monopolies in existence in one country, their control over prices will be limited by the ever present threat of foreign competition.

© ABE and RRC

We must recognise that the threat of competition is often weakened by the development of large multinational companies. Such companies tend to limit world competition by agreements between themselves, and by their own power to absorb competitors.

(e) Encouragement of International Cooperation

The existence of international trade also leads to a greater degree of interdependence between sovereign states, and this should be a factor making for international peace and friendly cooperation between nations.

Comparative Cost Advantage

In addition to these benefits, economic theory suggests a further benefit that enables us to explain why countries may buy goods which they could quite well produce for themselves. However, before examining the concept of comparative costs, we can consider an example where there are some fairly obvious gains from specialisation and trade.

Let us assume that there are only two countries, A and B, and that these countries produce only two commodities (disregarding any commodities which could not enter into international trade), which are wheat and copper.

Assume that, for a given outlay (which might be measured in terms of labour and money):

- A can produce 300 units of wheat and 150 units of copper
- B can produce 150 units of wheat and 100 units of copper.

Country A apparently has an advantage over country B in the production of both wheat and copper. Both commodities can be produced more cheaply in country A, as with a given outlay more of each will be produced in A than in B. Will there be any scope at all for international trade? The answer will be in the affirmative, provided that A's advantage over B *is not proportionately the same for both commodities*. A country will thus tend to specialise in the production of those commodities in which it has the *greatest comparative advantage*, or the *least comparative disadvantage*.

Let us now illustrate this principle with the help of our example. In the absence of international trade, A will produce 300 units of wheat and 150 units of copper, and for the same outlay, B will produce 150 units of wheat and 100 units of copper. This makes a total of 450 units of wheat and 250 units of copper. In A the cost of production of wheat is half that of copper, while in B it is two-thirds that of copper. As A's comparative advantage in the production of wheat is greater than its advantage over B in the production of copper, it will pay A to specialise in the growing of wheat and to leave copper production to B.

Suppose B abandons production of wheat and concentrates on copper: then A can make good the lost 150 units of wheat by transferring half the original outlay from copper to wheat. This still leaves A producing 75 units of copper, in addition to the increased 100 units of copper in B. Thus, specialisation in each country has increased copper production without any loss of wheat. Provided both countries trade with each other to share the increased production, both can gain from specialisation and trade – and A can gain by reducing its production of copper and importing from B, even though it is more efficient as a copper producer.

Table 15.1 illustrates the example just described. Here, the "given outlay in resources" is assumed to be 20 workers available for producing either commodity.

© ABE and RRC

	Country A		Country B		Total
Product	**Units**	**Workers employed**	**Units**	**Workers employed**	**Units**
(a) Before specialisation					
Wheat	300	10	150	10	450
Copper	150	10	100	10	250
		20		20	
(b) After specialisation					
Wheat	450	15	0	0	450
Copper	75	5	200	20	275
		20		20	

Table 15.1: Advantages of specialisation

The same total resources (40 workers) now produce an additional 25 units of copper, without any loss of wheat.

Limitations to the Gains from Comparative Advantage

It is sometimes argued that because of comparative advantage, there will always be gains from international trade, and that such trade should be freed as much as possible from government rules or restrictions. Before accepting this, we should remember that there can be general gains from increased specialisation and international trade only if:

- production factors, including workers, are able to move from one activity to another within each country – i.e. there is factor mobility within countries
- no factors are left unemployed and unproductive as a result of the movement resulting from increased specialisation
- there is a demand for the increased product made possible by changes
- there is no movement of production factors between countries.

For instance in the example just given, if the advantage of country A arises out of superior managerial skill, then the greatest gains might be achieved by exporting managers from A to B and improving the standard of production in B.

These are very important qualifications, and they do not always hold good under modern conditions. Production today is often highly specialised, and it is difficult – and sometimes impossible – to transfer resources (including workers) from one activity to another within a country. Machines are often built for one purpose only, people may take years to retrain, and unions are often hostile to movement. Many people displaced from one activity are just not able to learn the skills required for another (expanding) activity. In these circumstances, it is not unusual to find high unemployment in some sectors of production and a shortage of workers in another.

© ABE and RRC

B. TRADE AND MULTINATIONAL ENTERPRISE

The Multinational Company

The traditional theory of international trade based on the concept of comparative cost advantage now requires some reconsideration. There is no general agreement on a precise definition of a multinational company. For our purposes, we can regard it as any company which produces goods and/or services in several different countries. The company must own and directly control production facilities in the various countries. This is often referred to as "direct investment" overseas, and it is in contrast to "portfolio investment", where the home company simply owns shares or loan stock in foreign enterprises, and does not directly control their activities.

The term "multinational" usually conjures up an image of a very large company – indeed, the leading multinationals are giant enterprises. These include the oil producers and the mass-production motor manufacturers. On the other hand, there are many small companies which operate across national boundaries and take advantage of modern communications.

There are multinational companies owned and directed in many different countries, but the majority are American, European or Japanese owned. Until recent years Japanese companies preferred to concentrate production in Japan and export to the rest of the world, but as a result of several trends and pressures they have now started to take the multinational path to expansion.

Reasons for Growth of Multinational Enterprise

There have been some large world scale producers for a long time. The British Hudson Bay Company and the British and Dutch East India Companies were large organisations as early as the eighteenth century. However, these grew out of trading enterprises. Worldwide manufacturing is a development that belongs more to the twentieth and twenty-first centuries, and especially to the period after the Second World War.

There are many reasons for this development. Among those most commonly put forward are the following.

(a) Improvements in Transport and Communications

In a world of air travel and international telephone, fax and telex links, it was possible to retain control over the day-to-day activities of a worldwide enterprise in a way that would have been impossible in earlier times.

(b) Efficient International Capital Markets

An international banking system has developed with the growth of world trade and the spread of European influence in other continents. Bankers are often anxious to finance local branches of the large worldwide companies, sometimes in preference to more risky local business. Restrictions on capital movement from countries such as the USA and the UK in the 1950s and 1960s also tended to ensure that money earned in foreign countries was kept abroad to finance foreign direct investment, because if it returned home it was likely to be kept there by government controls.

(c) Encouragement by Developing Countries

The developing countries in Africa, Asia and South America offered growing markets for a wide range of goods. Many encouraged the entry of foreign manufacturing companies as a means of speeding up national industrial development and of earning much needed foreign currency from industrial exports.

© ABE and RRC

(d) Rising Costs and Production Difficulties in the Industrial Nations

Growing state intervention, the rise of trade union power and rapidly increasing wage, land and other production costs in the USA and Europe encouraged many companies to look to investment opportunities in developing countries. In such countries costs were lower, and there was much less resistance to the introduction of new machines and working methods.

Japanese companies have also been influenced by increasing production costs (especially wage costs) within Japan, and have established production divisions in other countries in both Asia and Europe.

(e) Product Life Cycle

If a company builds up a large export trade for a product, and if that trade is directed towards countries whose development is a little behind that of the home country, the time is likely to come when the export market in the developing countries is larger than the domestic market in the country of manufacture. By this time in the life of the product, it is probable that competition is developing from firms situated inside or closer to the export market, and the home market may also be starting to decline. It may well be that the production facilities will need replacing.

At this stage, the manufacturer is likely to consider setting up new production facilities (factories and machines) in the developing countries, where markets are growing. The remaining market at home can be fed from imports from the new factories.

In practice, some or all of these influences may be operating at the same time. The more influences that do bear on an industry, the greater the likelihood that it will become multinational in character.

(f) Trade Barriers

Some countries and groups of countries discourage imports by tariffs and other trade barriers. The European Union (EU) has established free trade between members, but it has many barriers to trade with non-members. It has been particularly restrictive against agricultural imports from developing countries.

Consequences of Multinational Enterprise

Multinational enterprise involves a transfer of production capacity from one country to another. It has consequences for the home country of the multinational company, the host countries where new enterprises are established, and for the whole pattern of international trade and production.

(a) Consequences for the Home Country

If a British manufacturing company decides to locate a new factory in Brazil rather than in England, then England loses the investment to Brazil. From the British point of view, this is called "divestment" – i.e. the loss of productive investment. The decision may mean a loss of some capital. However research indicates that much foreign investment takes place with the help of locally raised capital, and that the amount of finance exported, even to developing countries, is relatively small.

In the home country there is a loss of production work and jobs are lost. Most of these jobs are likely to be in the routine work of manufacturing – the unskilled and semi-skilled jobs and the work of supervision. The more highly skilled work of research, planning, marketing, etc. is still likely to be controlled by the home headquarters of the multinational company. Home country nationals are also likely to be asked to fill managerial and skilled technical jobs in the overseas country. It is possible that there are now more British people working overseas than there were in the days of the British Empire. The American and Japanese multinational companies are even more

© ABE and RRC

likely than the British to ensure that managerial and technical posts are filled by their own nationals.

Another consequence of divestment for the home country is that visible exports fall and visible imports rise. Invisible earnings rise, as the overseas sections of multinationals pay fees and royalties for patents and services, and remit profits to the home country. Of course profits go to the owners of capital – insurance firms and funds – and do little to make good the loss of jobs suffered by industrial workers. There is also a good chance that the profits will be reinvested in further foreign production, and not used to develop business at home.

(b) Consequences for the Host Country

The host country gains jobs and some capital investment. If local capital is raised, then this is denied to the country's own domestic industry and commerce. The country also gains export earnings and saves some import payments, by having producers of products for world markets within its own economy. There is some doubt whether it gains the full value of production though, because the home part of the multinational company will require heavy payments for technical and managerial services, as well as a substantial share of profits. It is notable that the group of what are now called the "newly industrialised countries" (Korea, Greece, Hong Kong, Mexico and others) nevertheless still have a balance of payments deficit with the advanced industrial countries. This is in spite of gaining a substantial share of world production of a growing number of industries (textiles, shoe manufacture, electronic equipment).

It is frequently claimed that host countries gain benefits from importing managerial skills and technical know-how. There is certainly some transfer of managerial skill and technology but this can be exaggerated, especially where the majority of skilled functions are kept for nationals of the home country, and where the home country retains full control over all research and development.

It will be in the interests of the multinationals to keep factor costs low, and for labour to be non-unionised. This means they will not encourage the development of domestic industries which may prove to be competitors, both in selling products and as employers of production factors. If factors (especially wage costs) do start to rise, then the multinational may be able to transfer production to another country, leaving the original host country worse off than before.

(c) Consequences for International Trade

There is no doubt that the growth of multinational enterprise has changed the pattern of international trade. Visible trade is no longer a matter of a flow of basic materials to the western industrialised countries and a counter-flow from them of manufactured goods. Manufacturing is now carried out in a very wide range of countries, though much of it is still controlled by and relies on technology supplied by the advanced industrial nations.

Even more important perhaps, is that the multinational companies have shown the importance of factor transfer between countries. Consider again the example of specialisation based on comparative advantage given earlier in this study unit. You will see that the whole process is transformed if we allow for the possibility that A's superiority in the production of both products is the result of superior managerial skill, and that this skill could be transferred from country A to country B. We cannot then predict the result of the transfer, because this will depend on which industries are affected, and on which terms the transfer takes place.

What we can say is that multinational enterprise on a large scale further undermines the theory of comparative cost advantage as the basis for international trade and exchange. Multinationals will locate in those areas where costs will be lowest for

© ABE and RRC

themselves in absolute terms. They are not concerned with the domestic comparative or opportunity costs of local factors they employ. They will seek that combination of local and "transported" factors (managerial skill and technology) which will give the production levels required at minimum cost. This is likely to mean that some parts of the production process will take place in one country and some in others. We can now see the association between the growth of multinational enterprise and the trade in semi-manufactures, much of which is intra-company trade – i.e. transfer between sections of the multinational companies.

C. FREE TRADE AND PROTECTION

Advantages of Free Trade

The principle of comparative advantage shows that free trade and specialisation brings gains to the participating countries. So long as a country has a comparative advantage in producing something it can benefit from specialising in its production, and trading the surplus over home consumption for other materials and products from abroad.

The advantages of free trade can be summarised as being that:

- countries can specialise and increase production safe in the knowledge that they can export their surplus
- resources are allocated efficiently
- countries can export surpluses and import what they need
- countries gain economies of scale from access to the world market
- competition from imports increases efficiency and limits the creation of monopolies
- free movement of capital allows countries to develop their industries
- political links develop between countries.

The factors of production are immobile. Land, most labour and invested capital cannot move between countries. Only enterprise, uncommitted capital and some labour can move to where the other factors are abundant and production can be organised. Free trade overcomes the immobility of factors: it permits the free movement of the product of immobile factors so that countries worldwide can benefit from an abundant factor endowment in any place.

Access to the global market is essential for developing countries if they are to achieve economic growth. Trade with the developed economies would give the developing nations a large market for their goods and the opportunity to import new technology. Firms could gain economies of scale and new techniques; competition would increase efficiency; monopolies are avoided. Production for export helps to diversify the economy: it reduces dependence on what is often a single crop subject to disasters, like sudden frosts which halve the output of coffee.

Protection

All trading nations engage in some form of trade protection, as governments have to face political pressures from powerful domestic interest groups. At the same time they are often reluctant to admit that they are imposing barriers, so they may avoid the formal measures that would invite retaliation and invite censure from the World Trade Organisation (WTO). Instead they make use of a variety of devices to delay imports or make them more expensive. These include cumbersome import procedures with complicated documentation or "safety measures" with a dubious safety value.

© ABE and RRC

At the same time the more formal measures still survive, and are employed by individual countries and regional groups such as the European Union. The main such measures are:

- *import tariffs*, also known as customs duties, which are taxes imposed on goods when they enter a country or one of a group of countries such as the EU, which contrast with
- *import quotas*, which are quantitative restrictions on the import of goods.

We examine these and other forms of protection in the next section of this study unit.

The belief that free trade (trade free from imposed restrictions) should be encouraged as much as possible is linked closely to the theory of comparative cost advantage. However, the benefits of comparative advantage have been shown to depend on the existence of competitive markets, absence of monopoly power, full employment, and ready factor transfer within countries and no factor transfer between countries. Instead of this, we have a world economy dominated by the monopoly or oligopolistic power of the large multinational enterprises. Few industries approach anywhere near the conditions of perfect competition, domestic economies are highly specialised, there is large-scale unemployment and little factor transfer within countries but important transfers between countries. In these conditions, we have to ask whether the case for free trade should be questioned and that for import controls looked at more seriously.

If a country does decide that, in its own case, the possible benefits of controls outweigh the dangers, the following arguments can be advanced in favour of the use of protectionist measures.

(a) Protection of "Infant" Industries

"Infant" industries need protection from foreign competition until they become strong enough to stand on their own feet. They are those industries which are being introduced to a country where the industry has not previously been present. The absence of external economies makes the costs of production high for new industries. In other countries, which are in competition with the country imposing the duties, the industries are already in existence and are therefore enjoying external economies of scale. As the infant industry grows, skills and productivity, as well as external economies, will grow also, so increasing the industry's relative competitive advantage.

Domestic pressures for protecting home industries are always greatest in periods of economic recession and high unemployment, as in the early years of the 1990s. There are also many people within the EU who would like to try and avoid the challenge of the emerging industrial nations of Asia, by erecting high barriers against the entry of goods from non-EU countries. On the whole, the opponents of increased trade protection have managed to contain the protectionist pressures, while the establishment of the WTO should ensure that these temptations will continue to be resisted and that further progress will be made towards reducing the present barriers.

(b) Protection against Dumping

It is sometimes suggested that measures are needed to protect a country against the dumping of foreign goods. "Dumping" means the application to international trade of the methods of a discriminating monopoly. Goods are sold abroad at a lower price than at home. This is done partly in order to avoid swamping the home market with a surfeit of goods which would bring down home prices, and partly in order to kill off foreign competition by undercutting it on its own markets. The alternative is "stockpiling", which means the goods may be released in times of need, or sold over a number of years under a controlled agreement. Dumping is generally looked upon as an unfair trading practice, and for that reason industries fearing competition from dumped goods ask for tariff protection.

© ABE and RRC

Here again some objections may be raised. The main objection is that many industrialists begin to complain if they have to face competition from foreign goods which are cheaper than their own. However this does not represent dumping if the exported goods are sold at the same prices at which they are available in their home markets. The home producers may simply be inefficient. Also, when dumping takes place, the imposition of protective duties may be too slow a weapon, since by the time the new duties have been introduced, the dumped goods may already be in the country.

(c) Increase in Employment

Controls cut imports and therefore there may be an increased demand for home-produced goods, and a resulting increase in employment. Income is directed away from foreign exports and towards domestic producers. On the other hand, if there is already full employment at home, such measures will tend to be inflationary in their effects.

(d) Improvement of the Terms of Trade

The imposition of import duties may lead to an improvement in the terms of trade, particularly where the goods taxed are in inelastic supply and elastic demand.

(e) National Security

Key industries, such as agriculture and those producing goods which are important for the defence of the country, must be maintained for security reasons. A wide diversity of industries is important to a country, as it renders it independent of foreign supplies which may be jeopardised in the event of war.

(f) Improvement of the Balance of Payments

This point has also been discussed already. However you should remember that the balance of payments is not only concerned with imports but also with exports, and the government will have to consider what effect the imposition of protectionist measures by a country will have on that country's exports.

(g) Possibility of Shifting the Burden

This is a hope which concerns any tax – i.e. that someone else will pay it. We have shown that this is likely to happen only if the foreign country's need to supply us is much greater than our own need to acquire that country's goods. This will be the case where foreign supply is inelastic – i.e. does not respond readily to price changes – while our demand for imported goods is elastic. If the higher price resulting from the imposition of import duties were to be passed on to the home consumer, purchases would drop substantially and the tendency would be to make up for the higher duty by reducing the import price of the commodity. If the price to the consumer in the importing country rises by less than the full amount of duty, the balance of the duty has in effect been borne by the exporter, in the form of a lower price received for the exported goods.

(h) Equalising the Costs of Production

It is sometimes suggested that competition from foreign producers who enjoy lower production costs is unfair, and that import duties should be levied at rates which would equalise costs, so that foreign and home producers would then compete on equal terms. This argument is quite nonsensical. International trade takes place just because there are comparative cost differences between different countries. If every country were to impose duties equal to existing cost differences, international trade would practically disappear.

© ABE and RRC

There is also a practical argument against the theory just outlined. Cost differences may refer to one of two things: they may refer to *basic* costs (i.e. differences in wage rates, rents or interest rates) or they may refer to *total* costs.

For instance, the fact that wages in a certain country are lower than in the United Kingdom does not necessarily mean that either wage costs or total costs in that country are lower than in the UK. It might be that labour is less efficient than UK labour, or it may be wastefully employed. Moreover labour is only one factor of production, and its productivity usually depends on both managerial skill and the availability of modern capital equipment. Countries with low wage costs are often short of capital, so that finance and equipment are frequently scarce and expensive. Countries with high wage costs, but with high levels of labour and managerial skills and ready access to capital, need to adopt different production methods from those applied in low wage cost countries.

Dangers of Trade Protection

The case against import controls is based on the following factors.

(a) Continued Faith in the Benefits of Free Trade Based on Comparative Cost Advantage

It can be argued that multinational enterprise, unemployment and specialised production represent modifications and imperfections only, and do not change the fundamental truth and importance of the benefits to be derived from international specialisation and trade. According to this view, efforts should be made to reduce the harmful effects of these – including efforts to reduce the monopoly power of large multinationals – and to increase trade, not to interfere with it.

(b) The Fear of International Retaliation

If all countries sought to reduce, and impose barriers against, imports, total trade and production would fall, and unemployment would increase in all countries. Far from being a cure for unemployment, the spread of protectionism would increase it.

(c) Reduction in Industrial Efficiency

Those who believe that competition is the main incentive to business efficiency fear that protecting domestic industry against foreign competition would make firms less able to compete in world markets. The longer controls lasted, the more they would be needed, and the country would lose the variety of products provided by imported goods. Its standard of living would fall with this loss of choice, as increasingly inefficient firms required more and more resources to produce less and less.

Those who favour import controls argue that the case for free trade based on comparative advantage has been weakened, as already explained. They also argue that controls are no more harmful to world trade than the other measures which have been used in the past to correct balance of payments deficits, and much less harmful to domestic production. They may even be less harmful than deflation and devaluation, because they can be more discriminating. Deflation harms all forms of production. Deflation also damages domestic industry by reducing total demand, and this tends to harm some industries more than others. When demand rises again, these industries may not be able to recover, with the result that imports rise to an even greater extent than before. Successive deflations produce ever-increasing imports.

Import controls can be applied more heavily in those industries where the home firms are weakest and, it is argued, help them to recover their lost markets. Where home industries have been completely lost or very seriously weakened by past policies, it is suggested that state aid may be necessary to bring about recovery. In these cases, continued protection would be needed until they were strong enough to compete again in world markets.

© ABE and RRC

Supporters of import controls argue that as the total effect is no worse than other measures to reduce balance of payments deficits, there is no reason why the danger of retaliation would be any greater. They also suggest that controls have the effect of reducing the propensity to import rather than the absolute level of imports, and so allow the economy to expand more readily. Any reduction in the marginal propensity to import will increase the value of the national income multiplier. An expanding economy could actually permit more total imports – rather than less – as a part of increased total consumption.

D. METHODS OF PROTECTION

A country which has nevertheless decided to restrict the freedom of international trade can use many methods. The main methods of protection are:

- tariffs (customs duties)
- quotas
- embargoes
- voluntary export restraint (VER)
- export subsidies and bounties
- non-tariff barriers applied through safety rules and administrative controls
- exchange control.

Tariffs

Tariffs – or customs duties – are taxes on imported goods and so of course they raise money for the government. The object is to raise the cost of the imported goods so that importers have to raise prices or accept reduced profits. The imports thus suffer a competitive disadvantage compared with home produced substitutes. The tariff raises the price paid for the imported good by the domestic consumer and reduces the quantity purchased. Thus domestic producers supply more to the market, and foreign suppliers provide less than if there were no tariff.

Customs duties may be imposed by a specific duty of so much per item or per tonne or ad valorem (by value). Specific duties work best for goods of low value and high weight, such as iron. Ad valorem duties obviously have more impact as goods increase in value, so they are best applied to items like jewellery and those whose prices change often.

The amount received by foreign exporters may be the same or less than before the tariff depending on the elasticity of demand. The more price elastic is the demand for the product, the more the producers have to absorb the effect of the tax to prevent a loss of sales which would cause them a loss. The effects of a tariff are shown in Figure 15.1.

© ABE and RRC

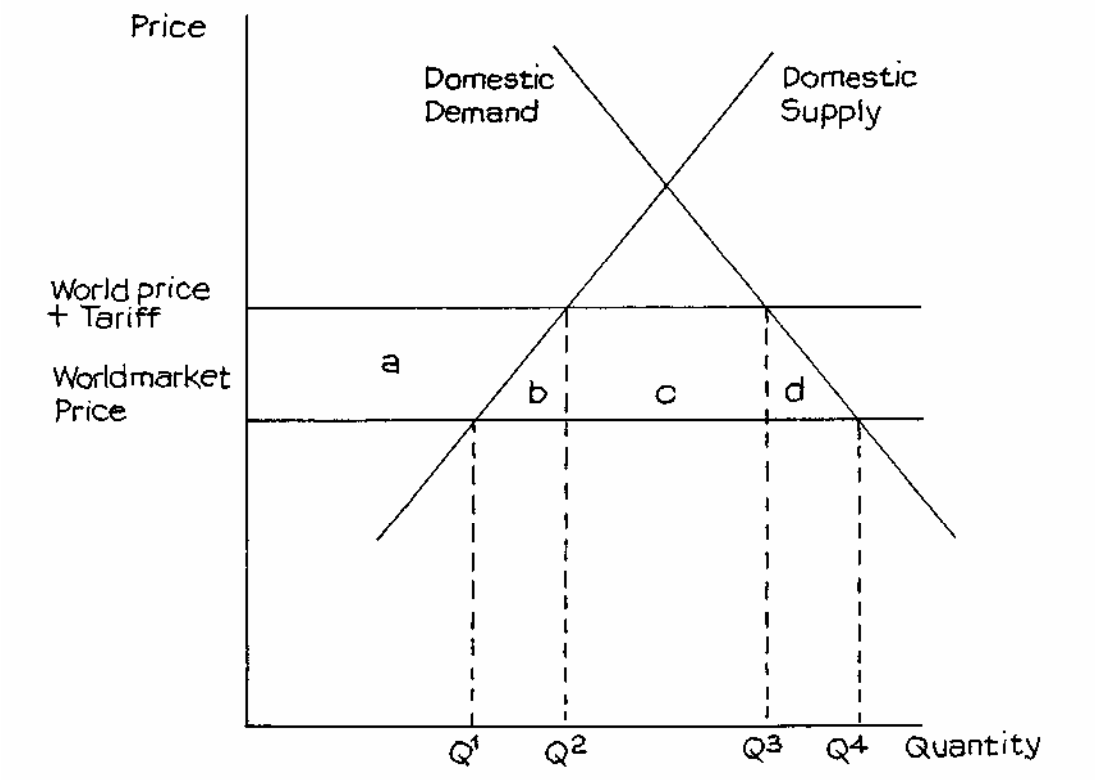


Figure 15.1: The effects of a tariff

The gross cost to consumers of the rise in price caused by a tariff is the sum of the areas

$$a + b + c + d$$

where:

- a represents a redistribution of income from consumers to producers
- b is the production cost arising from the misallocation of domestic resources
- c is the tariff revenue paid by consumers to the government, and
- d is the loss of consumption in the country imposing the tariff.

Areas b and d added together give the net costs of tariff protection to the economy. Tax and the additional domestic supply remain in the economy.

Not only do consumers pay a higher price and buy less, but there is also some loss of economic welfare because they are forced to buy the domestic product, which restricts their choice.

Quotas

Quotas are restrictions on the quantity of a product which can be imported.

While the purpose of protective customs duties is to restrict the import of goods by making them more expensive to the home consumer in order to persuade consumers not to buy them, the purpose of import quotas is to lay down the exact quantity of a commodity which may be imported in a given period of time. Import quotas may, but need not, be accompanied by customs duties. If they are, it means that the limited amount of goods which may be imported is subject to the duty as well. Quotas first came into prominence during the 1920s and 1930s, but they have also been widely used since the Second World War.

© ABE and RRC

The reasons why some countries prefer to substitute quotas for customs duties or to strengthen protective duties by quotas are as follows:

(a) Protective duties are sometimes considered to be insufficiently protective. This is particularly the case where the duty is a specific one rather than one related to the value of the imported goods. A specific duty is one which is imposed at so many pence (or pounds) per unit of commodity. At a time of quickly rising prices the specific duty becomes a declining proportion of the price of the commodity, and so loses much of its protective value. Frequent changes in the rate of duty may be difficult to administer, and would also lead to strong protests from the countries importing the goods. Thus a quota appears to provide the simplest solution to the problem.

(b) Quotas may generally be altered by administrative means – e.g. by an order by the Department of Trade and Industry. On the other hand customs duties are taxes, and as such they are subject to parliamentary control. If it is desired to strengthen or to relax protection, a change in customs duties might be hotly contested in Parliament, while a change in quotas could be brought into effect without much ado.

(c) Many pre-war international trade agreements expressly prohibited the participating countries from changing their existing customs duties, and the imposition of quotas was one way of getting round this restriction.

(d) Quotas also lend themselves admirably to a policy of discrimination. With customs duties, the same rate of duty will normally be payable on goods of a certain kind, irrespective of the country from which they come. A country wishing to reduce the volume of its imports may wish to cut down imports from a particular source – e.g. because the country concerned has a so-called hard currency, i.e. a currency which is in short supply. This end may be achieved by a quota scheme under which different countries are allocated different quotas, the quotas for goods from countries with soft currencies being rather more generous than those for countries with hard currencies.

(e) An occasionally heard (if mistaken) argument in favour of quotas is that quotas, unlike customs duties, will not lead to higher prices. This argument is wrong because, if a quota is effective in the sense that it lowers the supply of certain imported goods, these goods will then be in scarce supply in relation to the demand. This situation will inevitably lead to higher prices.

Embargoes

An embargo is a total ban on imports or exports, usually applied for political reasons. A recent example is the United Nations embargo on exports of armaments to Iraq and on oil exports from Iraq.

Voluntary Export Restraints

VERs are quotas operated by exporting countries. They are usually applied to avoid the more severe effects of government imposed tariffs and quotas. Thus Japanese car manufacturers operate a VER on car exports to the UK and the EU. A VER tends to prevent new firms from entering the export market. The permitted exports tend to be the more expensive versions of goods, as this earns the most profit from a restricted quantity.

Export Subsidies and Bounties

These can be of the visible type, where a bounty is paid to exporters by the government according to how much they send abroad. WTO rules generally forbid bounties, so hidden subsidies tend to be provided instead. For example, exporters get government insurance against political and commercial risks at very low rates, tax concessions on equipment used for making exports and help with borrowing to finance export production.

© ABE and RRC

Non-tariff Barriers

This is a term used to cover a multitude of measures applied to restrict imports, especially where countries cannot use tariffs and quotas because they belong to WTO or a free trade area. They include oppressive safety measures, like the USA requirement for destructive car tests, which would require the whole annual output of a small specialist manufacturer to be crashed. France attempted to keep out Far Eastern video recorders by insisting they went through one small, remote customs post where there were bound to be very long delays in clearing them. In the 1970s Britain required importers to pay an advance deposit on all goods: this imposed an extra borrowing cost and pushed up the price of imports. Around the same time the UK had two rates of VAT: the higher rate applied to goods like motorbikes which were mostly imported.

The term is also applied, when discussing trade liberalisation, to all restrictive measures except tariffs. This is because tariffs are the only measure to be visible and measurable with accuracy. Agreements to reduce tariffs are pointless if duties are replaced by other measures which are difficult to police.

Exchange Control

Control is enforced in many countries by requiring all buying and selling of foreign exchange to be done through the central bank; the currency is not convertible into other currencies of the holder's choice. The government can then allocate foreign exchange to whichever activities it considers should have priority. This is effectively the same as a quota and is subject to the same dangers. Governments can avoid some of the problems by auctioning off foreign exchange, as was done in Nigeria. The amount released to auction is determined by the state of the balance of payments. Governments have also set multiple exchange rates – for example the South African rand had a commercial and a financial rate until 1995 – and they can alter the value of the currency to make exports cheaper and imports dearer.

In recent years many governments have recognised economic damage done by exchange and capital controls, as well as their ineffectiveness in achieving what they were intended to achieve, and abolished them either completely or in large part. This is especially true of the world's developed countries and newly developed countries. The important exceptions amongst the world's rapidly developing countries in 2008 are China and India. However, both China and India have relaxed their controls, and indicated their intention to move to even greater freedom of currency and capital mobility.

E. INTERNATIONAL AGREEMENTS

Trading Blocs

Countries can join together in several different ways to obtain the benefits of free trade among themselves while keeping others out. What is included in the agreement depends on the political will of the members; they may be unwilling to expose agriculture to competition, or to accept the full degree of international specialisation which goes with completely free trade. Giving up some control of their national economies makes it difficult for countries to enter into these agreements.

There are effects on the direction of trade – some countries benefit and others lose. These blocs all have tariff walls which discriminate against imports from non-members. Trade may be diverted by the tariff from a low-cost producer country which is a non-member to a high-cost member state. The effects of trading blocs have to be carefully evaluated to see if they really do benefit the citizens of the member countries, and not just protect inefficient producers.

© ABE and RRC

(a) Types of Bloc

The types of international integration are as follows.

- In *preference areas* countries agree to levy reduced or preferential tariffs on imports from qualifying countries. The EU operates a system of preferences through its Association Convention, covering the former colonies of member countries.
- *Free trade areas* are where the members abolish tariffs on trade between themselves, but each country keeps its own tariff on imports from outside the area. This makes it necessary to have *rules of origin* to prevent imports being brought in through the lowest external tariff country. The North American Free Trade Area and the Association of South East Asian Nations are examples.
- *Customs unions* have free trade within the area with a common external tariff.
- *Common markets* are customs unions with additional measures to encourage the mobility of the factors of production and capital. The EU opened its common internal market on 1 January 1993. Citizens of the member countries can live and work anywhere in the EU, capital can move freely and there is a continuing programme of harmonisation of standards and regulations to permit the free flow of goods and services. The 1991 Maastricht Treaty agreed to a programme to move to economic and monetary union and to take the first steps towards political union by agreeing common foreign policies. Since 2003 the single European currency, the euro, has replaced the previous national currencies of the 15 member countries of the eurozone.

(b) Effects of a Bloc

Creating a trade bloc has two major effects:

- *Trade creation* **–** when a country which previously placed tariffs on imports from another member and produced the goods itself switches to buying such goods from another member country, this creates trade (although it may cause structural unemployment).
- *Trade diversion*, when the removal of barriers inside the bloc results in trade being switched from a more efficient producer outside the union to a less efficient one inside.

In addition to the benefits of trade creation, there are other benefits from setting up a free trade area:

- *Economies of scale* develop because the member countries now have a much larger "home" market.
- *Specialisation* in products having a comparative advantage creates greater opportunities of economies of scale.
- *Greater efficiency* is enforced because the members' industries are exposed to more competition.
- *Consumer welfare* is increased as people have more, better quality and cheaper goods, with more variety, to choose from.
- There is more *political cooperation* as the member countries develop common policies and become more dependent on each other.

Against this must be set loss of political and economic independence, because the countries must take into account the policies and rules of the bloc when deciding their own policies.

© ABE and RRC

The larger the trading bloc the greater the potential benefits, because of the better chance of including the lowest cost producer and the bigger opportunities for economies of scale. There will be more opportunities for trade creation, whereas there will have been a lot of duplication, and large cost differences, between the production of the members before the union. There will be more to be gained from specialisation. This is especially the case when there were high tariffs before the union; there would then have been a lot of domestic production for relatively small markets. The lower the external tariffs imposed by the union the better, as this reduces the possibilities of trade diversion.

(c) Monetary Union: the Single European Currency

As early as 1970 the (then) EEC had a plan and a programme aimed at achieving economic and monetary union by 1980. By 1974 the attempt had failed, although the development of the European Monetary System (EMS) in 1979 gave a new impetus to monetary union and, until its breakdown after 1992, the monetary discipline it imposed appeared to bring the economies of the Member States closer to convergence.

The Maastricht Treaty laid down rules and a timetable for monetary union through a series of stages, culminating in the establishment of a common currency and associated financial institutions and policies. The key stage was reached in 1998 with confirmation of the countries meeting the convergence criteria, and EMU started on 1 January 1999. The convergence criteria were that:

- planned or actual government budget deficits should not exceed three per cent of GDP at market prices
- the ratio of total government debt should not exceed 60 per cent of GDP at market prices
- one-year inflation rates must be within 1.5 per cent of the three best performing economies
- one-year long-term interest rates must be within two per cent of the three best performing economies
- currency of Member States must have remained within the narrow ERM band for the two previous years without devaluation.

Some softening of the requirements in the treaty, allowing for the debt ratio to be reducing and for the annual deficit to be ignored if it is temporary, enables more countries to meet the criteria. (Ironically, Britain – which has reserved the right to opt out and hold a referendum on future membership – is one of the few nations able to meet all the criteria.)

The European Central Bank, located in Germany, took over from the European Monetary Institute and became responsible for monetary policy as part of the European System of Central Banks (ESCB), the other members being the national central banks. The European Central Bank has to ensure that the ESCB carries out the tasks imposed on it by Maastricht, namely:

- to define and implement the monetary policy of the EU
- to conduct foreign exchange operations
- to hold and manage the foreign exchange reserves of the Member States of the EU
- to promote the smooth operation of the payments system for cross-border monetary transfers

© ABE and RRC

- to contribute to the smooth conduct of policies concerning prudential supervision of credit institutions
- to ensure the stability of the financial system.

The jury is still out on the success of European Monetary Union and the euro. The role and status of the euro on the world's money markets since its introduction as a full currency in 2003 has gradually improved, so that it now rivals the US dollar as a major international currency. It did not fare well in value against the US dollar over the early years of its existence, although its loss of value against other currencies made euroland highly competitive against other countries. However since 2005, it has risen in value against the US dollar and other major currencies.

There have been undoubted benefits to industry and commerce for the euro-using countries of the EU, with the problems and costs of doing business in two currencies disappearing. This has had the expected incentive and led to increased inter-regional trade between the euro-using countries.

The main unresolved policy debate has been over the implication of a single currency for fiscal policy, and the need to maintain fiscal discipline and integrate fiscal policies. This implies that countries have to give up much of their control of their individual economic policies. France, Italy and Germany have all broken the requirement for fiscal discipline and exceeded the maximum permitted figure for the ratio of government budget deficit to GDP. In addition several countries, especially France, have tried to compromise the independence of the European Central Bank by bringing pressure on it to relax its policy stance against inflation.

GATT/WTO and the Liberalisation of Trade

In 1944 the 23 countries which established the United Nations met at Bretton Woods. Their purpose was to set up three new bodies with the objective of improving the workings of the international economy after the war. These were the International Bank for Reconstruction and Development (the World Bank), the International Monetary Fund (IMF) and the International Trade Organisation. The first two of these were approved:

- The World Bank has funded major projects, social development and private enterprises in developing countries, by using the capital subscribed by the member countries as collateral for its borrowing.
- The IMF holds substantial resources, paid in members' subscriptions, which can be used to help countries with balance of payments difficulties. Its establishment represented an amazing transfer of sovereign powers by countries to an international body – during the period of fixed exchange rates up to 1972, it was given control of exchange rates.

However the International Trade Organisation was too much for the 23 countries to accept – they would not give up sovereign power over their trade. The result was the General Agreement on Tariffs and Trade (GATT), which has no controlling powers but has attempted to get countries to agree to liberalise trade through a series of conferences.

Trade liberalisation has been carried forward in a series of *GATT Rounds* (talks) which started in 1947 and reached the eighth (the Uruguay Round) in 1986. By that time, the average level of tariffs had been reduced from 40 per cent to 7 per cent. GATT had also had considerable success in ending trade discrimination, but several problems remained where major countries and groups had entrenched positions.

There are now over 100 members who agree to abide by the "most favoured nation" rule, which means that one member that grants trade concessions to another agrees to extend them to all members of GATT.

© ABE and RRC

Since it started in 1986, the Uruguay Round continued in a series of meetings, but by 1993 it had failed to make progress on certain vital areas. These included agricultural subsidies and protection for textiles, which are of interest to developing countries, and intellectual property (patents, etc.) and trade in services where the developed countries wanted protection.

However there was a last minute agreement in December 1993 which went far beyond anything which could have been expected in 1986. The new deal came into force in 1995, eliminating tariffs on 40 per cent of manufactured goods and reducing others substantially. Non-tariff barriers were also reduced and a new transparency in international protection established, as easy-to-hide non-tariff barriers were replaced by published tariffs. A new framework of rules on subsidies, trade restrictions and public purchases was agreed, agriculture was brought fully into GATT for the first time, and trade in intellectual property was also covered for the first time, giving protection to patents, copyright and trademarks. The French managed to exclude audio-visual services from the deal and the USA was unwilling to permit the inclusion of maritime services. Financial services were only partly liberated, with a reciprocity rule applying between countries, so that any liberalisation by one partner has to be matched by the other.

Despite these limitations, the agreement represents the largest ever liberalisation of trade and is expected to make the world $6 trillion wealthier – developed countries benefit from the removal of barriers to services, and developing countries benefit from freeing trade in agriculture and textiles.

For the longer term, the most significant development may have been the transformation of GATT into the new World Trade Organisation in 1993, with real powers to police protective practices. The WTO was immediately faced with a trade dispute between America and Japan over trading practices, and another between America and China over intellectual property, and has been dogged by disputes about the influence of developed countries and multinational companies, and under-representation of the interests of developing countries. This has meant that further trade liberalisation has been limited, although a major agreement on telecommunications was concluded in 1997. However, the most significant development since 1997 has been the granting of full membership of WTO to China, and the dramatic rise of China to become one of the world's leading exporters of manufactured goods.

© ABE and RRC

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. Explain the meaning of comparative advantage.
2. Explain the meaning of absolute advantage.
3. Outline the benefits of free trade.
4. What are the arguments that may be used to justify restrictions on trade between countries?
5. What is the difference between a tariff and a quota when used to restrict international trade?
6. A country that currently use tariffs and quotas to restrict international trade announces that it is going to abolish all barriers to international trade and allow completely free trade. Explain the possible economic benefits of the new policy if foreign firms decide to invest in the country by building new factories.

© ABE and RRC

Study Unit 16

National Product and International Trade

Contents ***Page***

A. International Trade and the Balance of Payments 286
Trade Revenues and the National Income 286
The Balance of Payments Accounts 290
Structure of the Accounts 290

B. Balance of Payments Problems, Surpluses and Deficits 293
Current Balance Surplus 293
Current Balance Deficit 294
Causes of a Persistent Current Balance Deficit 294

C. Balance of Payments Policy 297
Devaluation or Depreciation 297
Deflation 298
Import Controls 299
Need for a Healthy Business System 299

© ABE and RRC

Objectives

The aim of this unit, in conjunction with Study Unit 15, is to explain the fundamental advantages and disadvantages of free trade, including the principles of absolute and comparative advantage.

When you have completed this study unit you will be able to:

- explain how the various measures of the external account (for example, current account, capital account, balance on visible trade) are constructed
- describe the different factors which determine the state (surplus/deficit) of these accounts.

A. INTERNATIONAL TRADE AND THE BALANCE OF PAYMENTS

Trade Revenues and the National Income

We now return to our basic model of national income. Remember our proposition that:

total expenditure = total spending = total product

In a closed economy, where there are no foreign payment transactions (or where these are ignored):

- total income can be expressed as $Y = C + S + T$, and
- total expenditure, which also represents total demand (the desire to spend), can be expressed as $E = C + I + G$

where:

Y = national income

C = consumer spending

S = household saving

T = taxation

E = total spending

I = business investment and

G = government current capital spending

From these propositions, we saw that when national income and expenditure are in equilibrium – when total spending demand equals total production and income – then, because C features on both sides of the national income/expenditure identity,

$S + T = I + G$

If the government pursues a balanced-budget policy, then this will force savings towards equality with investment.

When we open up the economy to take into account foreign payment transactions, then this pattern has to be modified. If for simplicity we ignore non-trading transactions in international payments, then we can limit our consideration to the production of goods and services. Some of these will be produced at home and give rise to domestic factor incomes (exports), and others will be produced in other countries and bought at home (imports). Thus, some part of total income will be leaked away through spending on imports, while total spending demand will be augmented by the expenditure of foreign people on a country's exports.

© ABE and RRC

Imports are therefore a leak from the circular flow of economic activity, while exports can be regarded as an injection.

Using the symbol M for imports (because I has already been used for investment) and X for exports (because E has already been used for expenditure), we can now incorporate trading transactions into the model. We can do this either by adding to both sides of the equilibrium equation, i.e.

$$S + T + M = I + G + X$$

or we can emphasise the rather separate nature of these transactions by keeping M and X together. We can then ignore them on the income side and include them on the expenditure side, to produce:

$$C + S + T = C + I + G(X - M)$$

where $X - M$ represents the net expenditure flow resulting from the balance of trading transactions. If import payments exceed export receipts, then the net result is of course negative.

Notice that C has been reintroduced here, because we can regard much spending on imports as being a part of household consumption. Total import spending from total income will of course be made up of spending on consumer goods, investment goods, and goods required by the government.

If total imports equal total exports in value, then there is no direct effect on the size of the national income flow. Leaks are just balanced by injections. If import payments are greater than export receipts, then there is a contraction in the circular flow. If export payments are greater than import payments, then there is an increase. Remember always that it is payments that concern us, not volume. These effects can be illustrated as in Figures 16.1(a) and (b).

Here we see how a net excess of import payments brings down the equilibrium level of national income (Figure 16.1(a)), while a net excess of export earnings increases it (Figure 16.1(b)). This is what normal common sense leads us to expect. People gain jobs and earn incomes by providing and selling goods and services for export. On the other hand, if people spend their incomes on foreign-made goods, then this leads to the creation of jobs and incomes in foreign countries.

Another method of illustrating this is shown in the graphs of Figures 16.2(a) and (b). In Figure 16.2(a), we see the effect of increasing injections by net export earnings – the equilibrium level of national income rises from Oe to Oe_1.

In Figure 16.2(b), imports raise the slope of the withdrawals ($S + T + M$) curve to bring down the equilibrium income level (from Oe to Oe_1). Notice that net exports are shown as a parallel line, indicating that they do not rise directly as national income rises, whereas imports are shown as having a greater effect at higher income levels. This is because the consumption element in imports increases with higher incomes, showing a behaviour pattern similar to that of any other form of consumption.

© ABE and RRC

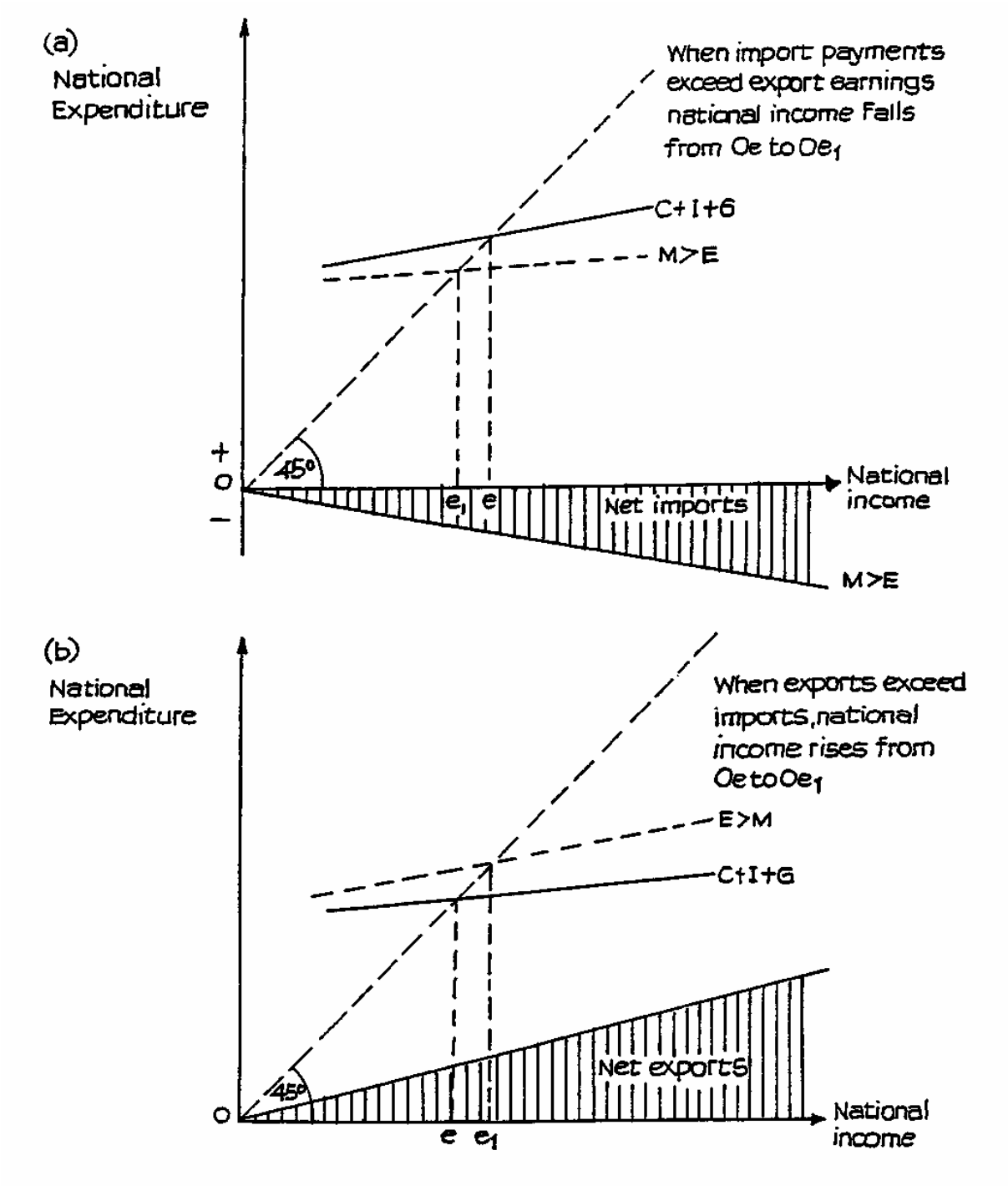


Figure 16.1: Changing the equilibrium level by imports and exports

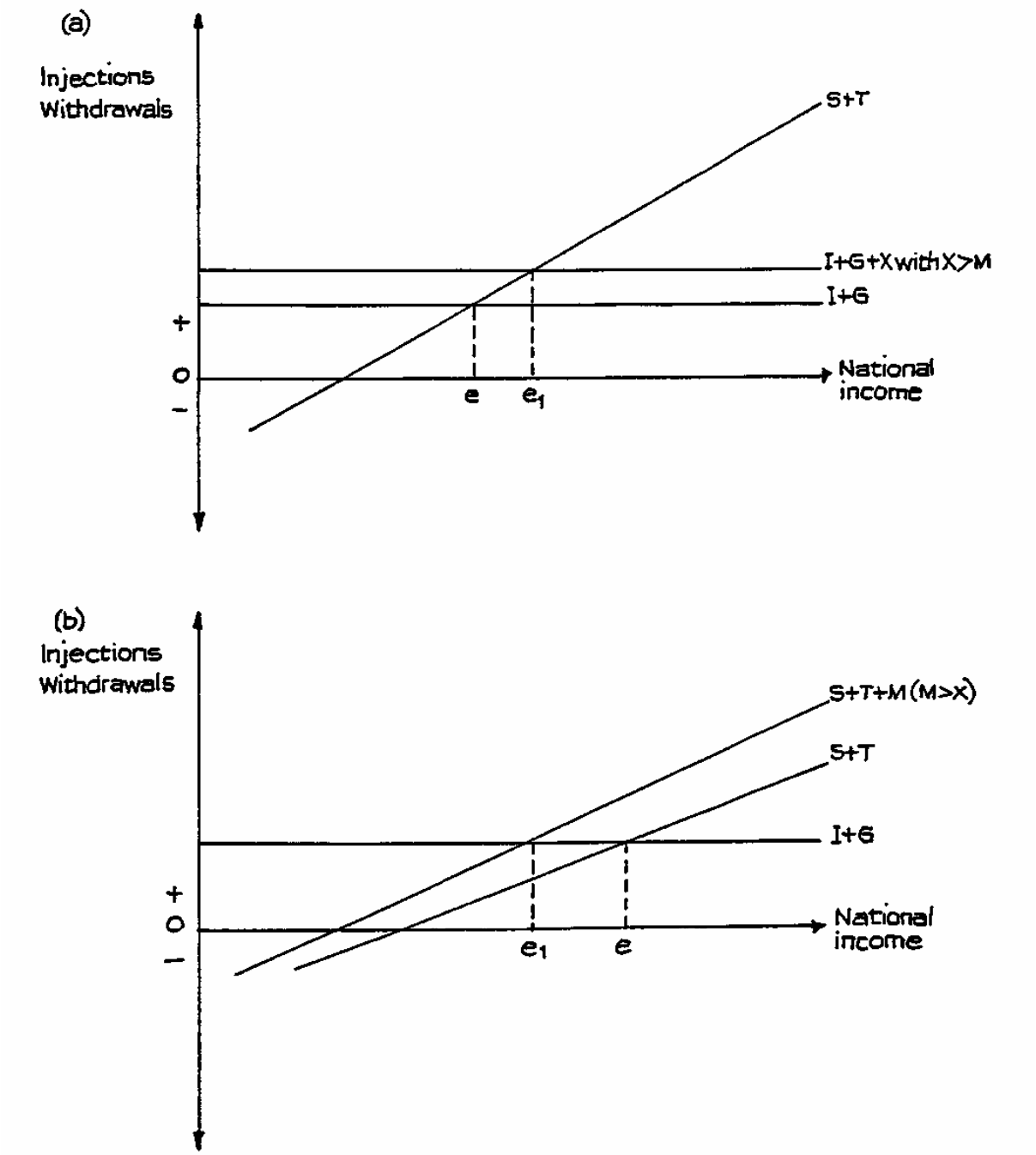


Figure 16.2: Another illustration of the effect of imports and exports

These illustrations help us to appreciate how imports reduce the value of the national income multiplier, in the sense that:

(a) increased consumption on imports makes the withdrawal curve steeper; the value of the multiplier is 1/w and any increase in the value of w, which represents the steepness of the withdrawal curve (the propensity to withdraw), reduces the value of the reciprocal of w;

(b) any increase in the import element in business investment spending reduces the net rise in I and, hence the injection brought about by I; if a firm buys machines made in another country, it is not creating jobs in home factories;

(c) any government spending on imports reduces the value of G to the domestic income in exactly the same way.

© ABE and RRC

There is nothing strange in any of these propositions. They are exactly what we would expect. However, we should remember that they all assume that the home and foreign economies are entirely distinct – i.e. that the home economy is not affected in any way by changes in foreign economies. A little further thought causes us to doubt this. Modern economies are closely interrelated.

It is true that there is no *direct* relationship between the size of the national income of country A and the level of exports to country B. However, if the two countries are trading partners, the national income of country B and its ability to buy goods from A will depend to some extent on its ability to sell its own products to A. There is a connection, and we should beware of making over-simple deductions from the apparently obvious propositions just given.

The Balance of Payments Accounts

The *balance of payments* is defined as a systematic record of all economic transactions between the residents of a country and the rest of the world during a period of time.

The national accounts which give details of payments and receipts and general financial transactions with other countries are called the *"balance of payments accounts"*. They mostly follow a fairly standard pattern, so that, although the following details relate chiefly to the United Kingdom, the main principles involved are likely to apply to most countries.

There are two main accounts:

- the balance of payments on current account
- transactions in external assets and liabilities (the capital account).

The current account is divided into:

- the visible trade account – the balance of trade
- the invisible trade account – services, transfers and interest, profits and dividends.

It is important to remember that the accounts represent flows of money. These flows are in the opposite direction to those of goods and services. For example, exports flow out, payment for them flows in; British ships carry goods for German firms and payment flows in. Capital investment by UK companies in America is an outflow of money, whereas the purchase by Americans of shares in British companies is an inflow.

Structure of the Accounts

The balance of payments accounts are shown in Table 16.1, where a minus sign represents money flowing out of the country and a plus sign indicates money flowing in. We shall then go on to discuss what the various items mean.

© ABE and RRC

		£ billion	
Current account			
Goods			
Exports		+121.4	
Imports	–134.6		
Balance of visible trade			–13.2
Services			
Exports		+36.6	
Imports	–31.6		
Interest, profits and dividends			
IPD receipts		+74.0	
IPD payments	–71.0		
Transfers			
Transfer receipts		+5.4	
Transfer payments	–10.5		
Balance of invisible trade			+2.9
Balance of payments on current account			–10.3
Transactions in external assets and liabilities (the Capital account)			
Direct and portfolio investment			
Investment overseas	–102.9		
Investment into the country		+49.5	
Net investment			–53.4
Bank transactions			
Lending abroad		+12.7	
Borrowing abroad		+23.7	
Net lending and borrowing			+36.4
General Government Transactions			
Overseas assets	–0.6		
Overseas liabilities	–0.1		
Net increase or decrease			–0.7
Domestic Non-banks transactions			
Lending overseas	–10.1		
Borrowing overseas		+12.7	
Net lending and borrowing			+2.6
Net transactions in assets and liabilities *(balance of payments on capital account)*			+8.3
Balancing item		+2.7	

(Note: The figures may not add because of rounding)

Table 16.1: The Balance of payments accounts

© ABE and RRC

(a) Visible Trade

When we think of trade, we usually think first of trade in actual physical goods, such as cars, oil, and food. This is normally called the trade in "visible goods", and the balance between the value of imports and exports is often called the "visibles balance". The correct term for this balance is the *trade balance* or *balance of trade*.

Visible trade is usually classified into a number of broad groups, and it is a useful exercise to look at the composition of UK trade on the basis of these groups. (You should try to obtain similar figures for your own country, if this is not the UK.) The main classes are the following:

- food, beverage and tobacco
- basic materials
- mineral fuels and lubricants
- semi-manufactured goods
- finished manufactured goods.

(b) Direction of Visible Trade Flows

You should also be aware of the main trading partners in this general process of international exchange. For example, Britain's main trading partner has, for some years, been the rest of the European Union (EU).

(c) Invisible Trade

Invisible trade is so called to distinguish it from trade in goods, which are tangible items. It consists of:

- *Services* including sea and air transport, tourism, consultancy and financial services.
- *Interest, profits and dividend* (IPD) comprises the annual flow of interest payments, profits from business and dividend payments on shares coming into a country from its lending and physical and financial investments overseas, less the payments of interest, profit and dividends due to foreign banks, companies and investors flowing out of the country.
- *Transfers* of funds to or receipts from other countries for non-trading and non-commercial transactions. The main source of transfers usually involves governments. For example, in the UK the government is responsible for most transfers in the form of grants to developing countries, subscriptions to international organisations like the United Nations and net payments to the EU. Private transfers include payments to dependants abroad by UK residents, and gifts.

The amount of IPD earnings depends on the amount invested in the past. Direct investment refers to the purchase of foreign assets. It includes buying control of firms in other countries, establishing subsidiaries and acquiring land and property. Portfolio investment is in stocks and shares. IPD receipts are influenced by the level of interest rates and the conditions in the economy which affect interest and dividend payments.

Profits and dividends in the balance of payments can cause confusion about how they appear in the accounts. If a British company has a wholly owned subsidiary overseas which earns a profit, the invisible earnings are the profit remitted to the UK. But the part of the profit which is retained in the overseas subsidiary is treated as a capital outflow, and appears under direct investments in the capital account. If the British company does not control the overseas subsidiary but receives a share of the profit, it only appears in the invisible account.

© ABE and RRC

(d) The Capital Account

The correct name for this account is *"transactions in external assets and liabilities"*. This account records only *changes* in assets and liabilities. For example, in the UK when the pound rises in value against other currencies, it becomes relatively cheap for British companies to invest abroad. Whereas if the dollar is strong compared to sterling, American investors will buy assets in the UK. Portfolio investment is undertaken by insurance companies, pension funds, unit trusts and investment trusts to diversify their portfolios and to seek gains from rising share prices in rapidly growing countries.

(e) The Balancing Item

The balancing item is a statistical adjustment to account for the failure to record some of the thousands of items in the current and capital accounts. It is the difference between the recorded entries in the balance of payments accounts and the change in official foreign exchange reserves.

Although people, the media and politicians talk about a country having a balance of payments deficit or surplus this is technically incorrect. When you hear or read about a country's balance of payments problem, usually it is a deficit or surplus on a country's current account that is being referred to. Because the balance of payments accounts are based on double-entry bookkeeping, the balance of payments of a country will *"always balance"*. However in effect this balance may have to be achieved by borrowing, from payments from past reserves and with the help of a balancing item which is often quite substantial! For example, if a country's balance of payments accounts show that it has imported far more goods and services in a year than it has exported to the rest of the world, it must also have already financed this deficit in some way unless the rest of the world has become very generous and supplied the goods and services for free! The really important balance though, is the current one. This shows whether the country is trading profitably and successfully or not. It is the current balance which is the best indication of a country's economic health. No country can overspend its current income and draw on past savings or borrow from other countries for ever.

B. BALANCE OF PAYMENTS PROBLEMS, SURPLUSES AND DEFICITS

Current Balance Surplus

We have seen how a surplus of revenue from all forms of export over payments for imports results in the equilibrium level of national income being raised.

The implications depend on whether or not the extra money available for spending pushes the national income equilibrium above the full employment level. If it does – i.e. if demand for goods and services is greater than the amount of goods and services available for purchase – then there will be inflation: prices will rise, or there will be shortages.

If it does not, then the extra inflow of money will generate extra economic activity, and unemployment will fall. The general level of employment and standard of living of the country will rise. This is often known as an "export-led boom".

However if there is pressure on the country's capacity to produce sufficient to meet the higher level of total demand, inflation may still be avoided. This is possible if the country exports some of the surplus money by investing abroad, or by making loans or grants to other countries. This may help these countries to develop their economies, and it will also help the revenue-exporting country's invisible balance in future years.

© ABE and RRC

The ability to allow or to encourage money to be used abroad will also help the country's political power and influence. It is little wonder that governments seek to achieve a balance of payments surplus on current account.

Current Balance Deficit

The equilibrium level of national income is reduced by an import surplus. In this case, money flows out of the country and the flow of goods and services in relation to the pressure is increased.

Again the immediate effect depends on the existing level of economic activity. If the economy is operating under inflationary conditions, with demand greater than can be satisfied at full employment, the deficit will reduce the inflationary pressure. People who cannot buy home-produced goods, because not enough are being made, will buy foreign goods instead.

However, if the economy is operating at lower than full employment, then the effect is to increase that rate of unemployment – more people will lose their jobs, and more machines will be idle.

In an advanced country, this can be only a fairly short-term effect, as a deficit causes other problems. These problems lead to measures to correct the deficit, and there are then yet further effects on the price level and on the extent of unemployment.

In a developing country, a deficit can be tolerated for a longer period, if it can be financed by foreign countries or by loans from the International Monetary Fund. This might be done as measures to raise general world living standards and increase the speed of world economic development.

In the advanced country, the outflow of funds to pay for imports will be greater than the inflow paid for exports. This means that the demand for foreign currencies to pay for foreign-produced goods and services is greater than the demand for the home currency to pay for that country's goods sold abroad. In this situation, the exchange value of the home currency is likely to fall.

Causes of a Persistent Current Balance Deficit

It is difficult to work out effective remedies for a balance of payments deficit, unless the causes of the problem are known. We have to admit that there is some uncertainty on this question. However it is possible to examine some of the influences operating on the pattern of a nation's trade.

(a) Changes in the Terms of Trade

The "terms of trade" measures the relative movement of import and export prices. It is calculated from:

$$\frac{\text{unit value index of exports}}{\text{unit value index of imports}} \times 100$$

The unit value index represents the average movement in price of a unit of imports or exports. The "unit" itself is a kind of average of all types of visible imports and exports. The terms of trade thus gives a general indication of how average import and export prices are moving. As an illustrative example, the actual calculations for the UK for 1983 to 1993 are shown in Table 16.2.

© ABE and RRC

		1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
Unit Value Indices (1990 = 100)												
Exports												
All goods	CGTO	86.3	93.1	98.1	88.4	91.5	92.4	96.6	100.0	101.4	103.5	114.8
Non-oil goods	CGSX	76.6	82.4	87.6	88.5	91.5	94.3	97.5	100.0	102.3	105.0	116.9
Imports												
All goods	CGTP	84.2	91.8	96.3	91.9	94.6	93.7	97.7	100.0	101.2	102.1	110.5
Non-oil goods	CGSY	80.0	87.3	91.7	92.1	94.8	95.0	98.2	100.0	101.7	102.9	111.8
Terms of trade												
All goods	CGTQ	102.5	101.4	101.9	96.2	96.7	98.6	98.9	100.0	100.2	101.4	103.9
Non-oil goods	CGSZ	95.8	94.4	95.5	96.1	96.5	99.3	99.3	100.0	100.6	102.0	104.6

Table 16.2 Terms of trade

We can analyse the results of changes illustrated by the index. At one time, a rise in the index was regarded as being favourable because a given quantity of higher-priced exports could earn enough to buy more imports. In the modern world, the results of trading-price movements are a little more complex.

- ***Import Prices Rise Faster than Export Prices***

  The effect will depend upon the elasticity of demand for imports. We can assume that in an advanced country, the demand for imported raw materials and foods and oil is fairly price inelastic. However the demand for most manufactured (especially consumer) goods is likely to be price elastic – provided that the home country is able to manufacture acceptable substitutes for foreign-made products. In this case, the demand for the price inelastic goods will fall in a smaller proportion than the rise in price, so that the total cost of payments for these imports will rise. In the case of imports the demand for which is price elastic, the fall in demand will be greater in proportion to the rise in price, and the total cost of these imports will fall.

  For a country such as Britain, where over half of the imports consist of manufactured goods, the effect of a change in import prices will depend on which imports are most affected. A price increase on foods, basic materials or imported oil would create a balance of payments deficit or make an existing deficit worse. If it is the prices of the manufactured goods that rise, we would expect there to be a fall in the total cost of imports. That is of course if demand is price elastic. If in fact there are not sufficient home-produced alternatives to make good the higher-priced imported products, then the demand may turn out to be inelastic and upset the predictions relating to total revenue.

  For a developing country, most imports are likely to be demand inelastic if they are needed to promote development, so that a rise in import prices would make for a deficit or aggravate an existing deficit.

- ***Rise in Export Prices***

  Again, the effect depends on the price elasticity of demand for exports. In this connection, a developing country exporting basic materials with price inelastic demand would gain, and would receive an increase in total export earnings. In a

© ABE and RRC

developing country, it might be difficult to absorb a large balance of payments surplus, and much of it might have to be invested abroad until the home economy could be developed. This was the case of some oil exporting countries when they gained from oil price rises.

One problem for a developing country that relies on the export of a few basic commodities is that its living standards are very much at the mercy of world prices of these commodities. When prices are high, the country might develop a standard of living highly dependent on imports, and this might be very difficult to maintain if world prices of the exported goods fall. It would be no use trying to stimulate demand by reducing prices, because this would only cut export earnings still further.

For a country such as Britain, chiefly exporting manufactured goods, export demand is likely to be price elastic and a price rise – caused, perhaps, by home inflation – is likely to lead to a fall in total export earnings, and hence to a deficit or the worsening of an existing deficit.

(b) Economic Weakness

Many economists think that relative price movements are little more than a symptom of economic conditions, rather than a basic cause of those conditions. For a developing country, a balance of payments deficit may simply reflect the world market situation that ensures that total export earnings for the volume of goods exported are not sufficient to provide enough money to pay for the goods and services needed for development. The position will be made worse if:

- world demand is declining for the country's basic exports, or
- there is a failure in production, resulting from natural disaster or other causes – e.g. a crop failure or internal conflict, or
- there is a high demand for imported consumer goods from a section of the population that has developed a fashion or taste for imported clothes, cars or food.

For an advanced country, the problem may be caused by a weak economic or business structure, an economy that is less successful than that of competing nations. If production is cut by poor working methods, under-investment in modern machinery or labour disputes, then export earnings are likely to fall and imports and the cost of imports rise, almost regardless of price advances, in favour of the home country. For example Germany and Japan have been consistently more successful in exporting than Britain and the USA.

(c) Activities of Multinational Companies

About a third of international trade is made up of payments between the different parts of multinational empires. These companies, operating on a world scale, may prefer to move production away from high-cost, highly-taxed and closely-regulated countries to other areas where they have lower costs and more control over production methods. It is notable that countries with a high proportion of multinationals – the USA and Britain – tend to have persistent problems with their balance of payments. On the other hand Germany and Japan, which until recently have not produced worldwide enterprises, have had very successful export records and few balance of payments difficulties. It will be interesting to see which effect the development of German and Japanese multinationals has on those countries' payments balances.

© ABE and RRC

C. BALANCE OF PAYMENTS POLICY

There are three main remedies for a balance of payments deficit. These are devaluation (depreciation), deflation and import controls.

Devaluation or Depreciation

By devaluation or depreciation we mean the reduction in the exchange value of a nation's currency in terms of foreign currencies. For example, before devaluation a British pound might be equal to US$2, but after devaluation it may be equal to only US$1.5.

If a country allows its currency to float on the foreign exchange market, then the value of its currency will fall if demand for the currency falls. For example if the demand for pounds falls and that for US dollars rises, the price of the pound is likely to fall relative to that of the US dollar. This is called depreciation, and is a normal part of the operation of foreign exchange markets.

Devaluation happens when a country operates a fixed exchange rate policy (see Study Unit 17), and the government decides to reduce the fixed value. The government can then simply change the value by declaration.

In whichever way it is brought about, a depreciation/devaluation raises the price of imports and reduces the price of exports, at least in the short term.

It is important to understand the distinction between devaluation (action by governments when exchange rates are fixed) and depreciation (fall in value of a currency as a result of market movements). But you must also recognise that governments do intervene in currency markets to try to influence market movements, and a change in interest rates is sometimes brought about by a government in a deliberate attempt to change the currency value.

The J Curve

It is sometimes pointed out that in the very short term firms cannot change their plans. It takes a little time for traders to react to international price changes resulting from exchange rate movements. Consequently, a swift devaluation or depreciation will increase the prices of imports and decrease those of exports without changing quantities traded to any great extent. The immediate effect of the price changes will be to deepen the balance of payments deficit. However fairly soon plans and trading patterns are modified, and we would expect demand for imports to fall and foreign demand for exports to rise. The result would be to reduce the deficit and, if the reactions were strong enough, to turn it into a surplus. This is illustrated by what is usually known as the J curve, as illustrated in Figure 16.3.

Importance of Demand Elasticities

For the changed trading pattern to replace a balance of payments deficit by a surplus, the rise in demand for exports at the reduced world price must increase export revenues by a greater amount than any increase in import costs resulting from the import price rise. It will of course help if the import costs actually fall. The desired gain in net revenues can only come about if the combined price elasticities of demand for exports and imports add up to a value that is more negative than –1.

Effect in Industrial and Developing Countries

In the case of a developed country such as the UK, where manufactured goods dominate exports and form a high proportion of imports, we would expect a devaluation to have a favourable effect on the balance of payments in the short term.

© ABE and RRC

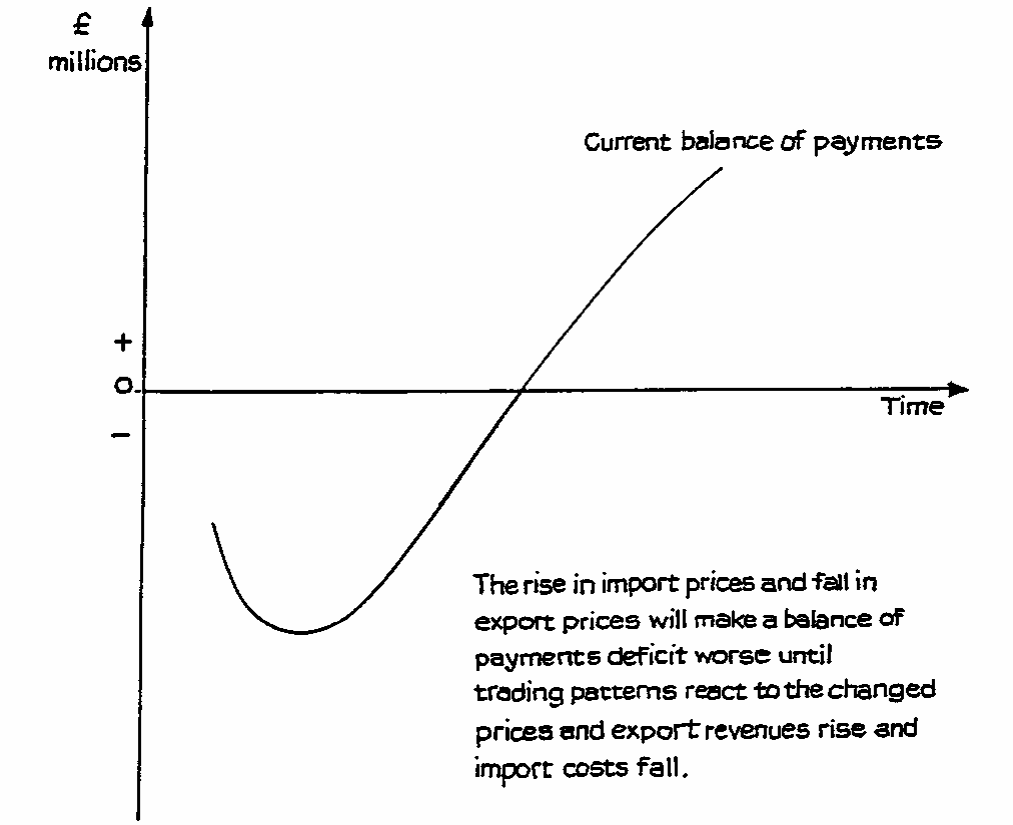


Figure 16.3: The J curve

In the long term, this beneficial effect of increasing net earnings is likely to be weakened. Any rise in the prices of imported fuels, raw materials and foods must soon increase the costs of manufacturing. It will also lead to an increase in the living costs of the workers. If the workers are able to secure wage increases in an attempt to restore living standards, then manufacturing costs will again rise. Inflation of both prices and wages thus erodes the competitive price advantages gained for exports against imports by the devaluation. If inflation continues at a high rate, the export price advantage may be lost very quickly.

For a developing country, both exports and imports are likely to be price inelastic. Thus the result of a devaluation in this case is to worsen an existing balance of payments deficit. The devaluation will reduce total export earnings and increase total import costs. Therefore devaluation will not help a developing country with balance of payments problems. It may help an advanced industrial country, but probably only in the short term. In itself, devaluation does nothing to cure the basic economic weakness which gave rise to the trading imbalance in the first place.

Deflation

Spending on imports is a form of consumption that is usually regarded as being dependent on the level of income of a community. The higher the income, the more is likely to be spent on imports. So one way to correct a balance of payments deficit is to reduce import levels or, at least to stop them rising too fast. A government faced with a balance of payments problem may seek to reduce disposable income in the hands of consumers, and so reduce all consumption expenditure. This will cut the demand for imports and also reduce the strength of demand for home-produced goods, so releasing them for export markets – if firms can be persuaded to make a bigger export effort. The government will achieve deflation by:

- reducing its own spending and the demand for workers in the public sector
- increasing taxes, and so reducing consumers' disposable (after-tax) incomes

© ABE and RRC

- increasing interest rates by restricting the money supply, so making it difficult for firms and households to maintain investment and consumption expenditure.

For a developing country deflation is unlikely to be a satisfactory solution, because the imports are needed for economic development. Also, if living standards are already very low, any reduction could lead to violent social and political unrest.

Import Controls

Countries can also attempt to remedy a persistent current account deficit by introducing control over imports through measures such as quotas and tariffs.

Supporters of controls suggest that the danger of retaliation is not as great as is often assumed, and they say that only with the protection of controls can the economy be fully revised. They usually also suggest that massive government aid would be needed for industrial modernisation and investment, and that the government would have to have greater controls over industry if it were to provide this aid. Taxes would also be likely to stay high if this policy were adopted.

Other people remember that it was the attempt of individual countries to impose controls over imports, and at the same time keep on exporting, that led to the trade wars of the earlier part of the twentieth century. These in turn helped to bring about the very severe depression and unemployment in the 1930s. They feel that the risk of such a tragedy being repeated is too great to allow import controls to be tried. However, the demand for controls is very strong in the face of what are often termed "unfair trading practices" of some countries.

Another danger is that industries do not in fact reorganise behind the protective barrier, and simply become less competitive and rely on satisfactory home demand. This is why advocates of import controls also tend to advocate increased public control to force modernisation.

The demand for import controls always increases during an economic recession, when there tends to be strong political pressure from industries with high unemployment rates or suffering from economic change to be given protection from foreign competition. There was a tendency in the late 1980s and early 1990s for informal methods of protection – the use of various administrative devices to make importing more difficult and expensive – to increase. The then GATT (General Agreement on Tariffs and Trade) negotiations for reducing tariff and other barriers in order to encourage world trade (originally due to be completed in 1992) encountered many difficulties, as governments sought to defend their own politically powerful groups – including of course the farmers.

The negotiations were eventually concluded by the end of 1994, and some progress was made towards further trade liberalisation. However progress was extremely modest in relation to the three major trading blocs of the EU, North America and Japan. At the beginning of 1995, GATT was replaced by a more structured body, the World Trade Organisation (WTO), which was given limited powers to enforce agreements and discourage openly protectionist measures. These were quickly tested by a trading dispute between the USA and Japan, though this was resolved without breaching WTO rules.

Need for a Healthy Business System

A balance of trade deficit for an advanced industrial country is a sign of economic weakness, and the only really effective long-term remedy is to strengthen the country's business structure. This means increased investment and business modernisation. This helps to explain why much more attention is given by governments than previously to the use of supply-side economic policy. Demand management policies alone are incapable of providing a lasting solution to balance of trade problems. The causes of a country's economic weakness in the face of stronger foreign competition are not always fully understood. They may be social or political, as much as economic. Devaluation, deflation and import controls

© ABE and RRC

are only short-term remedies. All may aggravate the weakness if no healthy business system is encouraged. There is unlikely to be a quick and easy solution, and some reduction in living standards may be inevitable before economic health is restored.

Review Points

Before you begin your study of the next unit you should go back to the start of this one and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. All other things remaining unchanged, how will an increase in the propensity to import affect the equilibrium level of national income of a country?
2. All other things remaining unchanged, how will an increase in foreign demand for a country's exports affect the position of its aggregate demand curve and its equilibrium level of national income?
3. Explain the difference between the current and capital accounts of the balance of payments.
4. If the balance of payments account must always balance explain the different ways in which a country can finance a deficit on its current account.
5. List the benefits to a country of allowing foreign direct investment into the country.

© ABE and RRC

Study Unit 17

Foreign Exchange

Contents ***Page***

A. International Money **302**
The Need for International Money 302
Gold – its Use and Limitations 302
Uses of National Currencies 303

B. Exchange Rates and Exchange Rate Systems **304**
What are Exchange Rates? 304
Effect of Exchange Rate Changes 304
The Formation of Exchange Rates 305
The Purchasing Power Parity Theory 305
Exchange Rate Structures 306

C. Exchange Rate Policy **310**

D. Macroeconomic Policy in Open Economy **311**

© ABE and RRC

Objectives

The aim of this unit is to explain how exchange rates are determined and to evaluate the relative merits of fixed and floating exchange rate regimes.

When you have completed this study unit you will be able to:

- explain the differences between the key terms used in the analysis of exchange rates: devaluation, depreciation, revaluation and appreciation
- explain the terms of trade
- examine the concept of purchasing power parity theory and its implications
- identify the relationship between fiscal/monetary policy and fixed/floating exchange rates
- explain the ways in which government manipulation of exchange rates can generate a competitive advantage.

A. INTERNATIONAL MONEY

The Need for International Money

We have seen earlier in the course that anything can serve as money, *as long as it is accepted as money.* It will be accepted only as long as it can be readily used to purchase real goods and services. Therefore money ceases to have any value as money when it cannot be easily traded for goods.

So the area of acceptability is extremely important for the value of any form of money, and this is a point of very great concern for matters of international finance and trade.

Therefore when one country sells goods to another, it wishes to be paid in a form of money (currency) which it can readily use to purchase its own goods elsewhere, or which it can change into its own currency to pay its own workers and suppliers at home.

You might think that it would all be a lot simpler if every country in the world used exactly the same currency, which would then be universal, and which would not be identified with any one nation.

Gold – its Use and Limitations

In a sense, there is a form of money which is universally acceptable and which is not associated with any one nation. This is gold, which has been used as money in almost every part of the world since the dawn of civilisation. Gold has all the qualities required of money. It is noticeable that whenever a country's financial or political system seems to be in a state of collapse, those able to do so abandon paper money in favour of gold which, if they can take it with them to another country, is readily acceptable there. Some international trading contracts are also arranged in terms of gold, and most countries keep at least part of their reserves in gold, the world price of which is a fairly good indicator of the general state of political tension in the world.

However, there is just not enough gold to meet the entire world's trading needs, and the natural supply of gold is very unevenly distributed between countries. If gold were the only international form of money, those countries where gold is found would have a degree of political power that other countries would find unacceptable. Moreover because gold, as a physical good, is in fixed supply in any given period, any of the metal that is held in reserve is withdrawn from circulation – and, thus, it cannot be used in exchange. Some countries, such as the USA, have such a large share of total world supplies in their reserves that they can influence its price (value in exchange for goods) by their sales in world markets.

© ABE and RRC

Gold – and, indeed, any other precious metal – does not provide an easy solution to the problems of international currency.

Uses of National Currencies

An attempt has been made to produce a form of "paper gold", to serve as a genuinely international currency. This resulted in the *"special drawing rights"* (SDRs) produced by the International Monetary Fund. But it has been found difficult to reach agreement on the issue and control of SDRs, and they have only a limited use in exchange and as a reserve.

The problem with any form of international currency is that there must also be some system of international control which all countries will accept. This immediately introduces political implications which so far have proved impossible to reconcile.

Consequently, the great mass of world trade has to be conducted in the normal national currencies of the world. Some of these are more acceptable than others, chiefly because some countries have stronger economies than others, and some governments have firmer control over their national economic and financial systems than others. For simplicity, we can identify four classes of currency used in international trade.

(a) The United States Dollar

The US dollar is the most widely acceptable currency, and it is used throughout the world. Many of the world's commodities and services are valued in dollars. They include oil and hotel charges. Dollars are also widely used in the internal trade of many countries, whose own currencies are very weak because of severe domestic inflation.

(b) Other Major Trading Currencies

The currencies of many of the other leading trading nations of the world have a wide acceptability, though not as universal and general as the US dollar. When the dollar itself is under pressure and losing some of its exchange value, one or more of these currencies becomes a refuge for international finance. Among the main trading and reserve currencies in this group are the euro, the Japanese yen, the British pound sterling, and the Swiss franc.

(c) Currencies with Limited Acceptability

Some currencies may be acceptable within a particular region. There are also many currencies, especially those of African countries and those of North Korea and Myanmar/Burma, that have almost no circulation or acceptability outside the national boundaries (and often are not too popular within the country either!). Sometimes a national government discourages international exchange involving its currency, as a means of keeping greater control and preventing the export of wealth. In other cases, the currency is too weak to support any external trade, or the official value in exchange for other currencies maintained by the national government is so unrealistic that no one who can possibly avoid it is willing to exchange foreign money at that rate.

(d) The "Basket" Currencies

These are currencies which are not the currencies of any nation, but their exchange value is based on a weighted basket of those currencies with which they are associated. The weights relate to the relative use of the various currencies for purposes of trade and international finance.

The main basket currency now is SDRs issued by the International Monetary Fund, although previously the ecu (European currency unit) was the basis for certain transactions within the (old) European Monetary System.

One of the advantages of using such a currency as a basis for valuing trading transactions, even if actual payments are made in a national currency, is that the

© ABE and RRC

basket currency fluctuates much less than any one of the individual national currencies. This is because changes in its value are simply the weighted average of all changes among the underlying currencies. Some of these are likely to cancel each other out: a falling currency could be balanced by a rising one. At present use of a basket currency for business trade and settlement purposes is restricted by lack of general availability, and also by lack of any widespread awareness of the position. People generally feel happier to stay with a currency they know and understand.

Trade may often be conducted by *barter arrangements* with some countries with weak currencies. For these agreements, some form of acceptable valuation is necessary. Again the basis of this tends to be the United States dollar, either directly or indirectly (e.g. through oil).

B. EXCHANGE RATES AND EXCHANGE RATE SYSTEMS

What are Exchange Rates?

We have seen that various national currencies are used in international trade, and we must now examine a little more closely what is involved when one currency is exchanged for another. The *exchange rate* is the rate at which the national currency can be exchanged for the currencies of other countries. Therefore there is not one rate but many, relating to all the different countries in the world. Some of the leading rates are shown in those banks which have a bureau de change (i.e. which can provide an over-the-counter service for changing currencies).

The principal rate which is of interest to most countries is the one relating to the main currency in use in international trade, the US dollar. For this reason we will concentrate on the US dollar/British pound relationship. For example, if the exchange rate is:

$1.20 = £1

then £1 can be exchanged for $1.20 (ignoring dealing and other costs of exchange). Thus:

£100 = $120

If however the rate changes to $1.10, then £100 becomes worth only $110.

Effect of Exchange Rate Changes

Suppose there is a fall in the value of the pound in terms of US dollars, so that in the space of a few months, the rate falls from $1.30 to $1.10. There is then an immediate effect on the prices at which traders are prepared to trade in international markets.

Say a manufacturer is prepared to sell a motor vehicle provided they receive £5,000. At the rate of $1.30 (again ignoring transactions costs), the manufacturer could sell the car in the USA for $6,500 (5,000 × 1.30). Suppose the pound falls in value and is worth only $1.10. Now the manufacturer will accept $5,500 (5,000 × 1.10) if they still wish to receive £5,000 for the car. Thus a fall in the currency value makes exports cheaper in foreign prices. Cheaper goods are likely to be easier to sell and, provided the increase in sales is proportionately more than the change in dollar price, exporters can hope to receive more *revenue* for their exports – hence, the use of devaluation to help in correcting a balance of payments deficit.

On the other hand imports become dearer, and this will affect the pound price of goods imported from other countries. Suppose the vehicle manufacturer buys steel from abroad and pays for it in US dollars. Each $1,000 worth of steel, which used to cost £769.23 (1,000 ÷ 1.3), now costs £909.09 (1,000 ÷ 1.1). Most manufactured goods contain materials imported from other countries, so that manufacturing costs inevitably rise following a fall in the exchange rate.

© ABE and RRC

There will also be other effects. A high proportion of British food and many consumer goods come from overseas – and so they rise in price. Living costs are pushed up and workers seek wage increases in order to try to maintain their living standards. If they succeed, then labour costs rise, and also manufacturing costs – and prices are also likely to rise. Under circumstances such as these, it is highly unlikely that manufacturers will reduce their foreign prices by as much as the full fall in currency value. In our example, the motor manufacturer will want more than £5,000. We can see that the effects of currency changes are far-reaching, and not always too certain.

The Formation of Exchange Rates

The exchange rate represents the price of the national currency and, like any other price; it is formed ultimately by the forces of supply and demand. These in turn are the result of the trade flows of imports and exports. In order to pay for imports priced in US dollars, the United Kingdom has to earn dollars by selling British goods and services to other countries. The more Britain can export, then the more dollars the country earns.

However British firms want to receive their payments in pounds. To obtain pounds to pay for British goods and services, foreign firms have to sell their own currencies in the markets for foreign exchange and buy pounds. So the greater the demand for British products in world markets, the higher is the demand for pounds in the currency exchanges. Conversely, the higher the demand in Britain for foreign products, the more pounds have to be sold to obtain the foreign currencies needed to pay for them. It is evident that one immediate cause of a change in currency exchange rates is the way the balance of payments is changing. If the balance is in surplus, then revenue from exports is greater than that paid for imports, and the supply of foreign pounds is high. So the pound is likely to rise in exchange value. A persistent balance of payments deficit has exactly the reverse result. The weaker the balance of payments, the weaker the pound is likely to be.

The views of traders and bankers about future movements in trade flows and currency exchange rates will also have an effect. For instance, traders often have to hold large sums of money for a few days or weeks, in anticipation of having to make large payments. They cannot afford to have money lying idle, so they lend it out in return for interest. They do not want to see the interest earned being lost through a fall in the exchange value of their money. This means that any suspicions that the pound is likely to fall will persuade the traders that their money is more safely kept in some other currency. This reduces the demand for pounds and increases the demand for foreign currencies, and so adds to the pressure resulting from a weak balance of payments. (Unless, as did the UK in 1989–91, the government tries to maintain an artificially high exchange rate through forcing up interest rates in order to attract sufficient foreign capital into the country to counterbalance the outflow of funds paid for imports.)

The Purchasing Power Parity Theory

If the immediate cause of exchange rate changes is a change in the flow of trade, then we are forced to ask whether it is possible to identify influences on these trade flows.

Various attempts have been made to explain these, and one such attempt is based on the view that they are directly linked to changes in inflation rates – i.e. in the relative purchasing power of the various national currencies. This is often referred to as the *"purchasing power parity theory"*. This theory states that the percentage depreciation of the home currency against a particular foreign currency can be expected to be equal to the excess of the home rate of price inflation over the other country's rate of price inflation. In other words, it is held that changes in currency values reflect changes in the purchasing power of the various national currencies. If country A has a higher rate of inflation than country B, then its currency buys fewer goods, and consequently it will fall in exchange value in terms of the currency of country B. This will continue until B's currency returns to the position where it will

© ABE and RRC

purchase roughly the same quantity of goods in A, when converted to A's currency, as it did before the price inflation. The theory is attractive but it is not entirely supported by the available evidence. It fails to take into account elements other than price which affect the demand for exports and imports. The theory also assumes perfect markets in currencies, but in practice governments tend to intervene to defend exchange rates. Governments can influence the rate of interest offered to investors or depositors of money. Traders may be persuaded to leave funds in London in pounds, in order to earn high interest rates likely to more than compensate for any change in exchange value.

In the long term, currency movements are most probably influenced by relative rates of inflation; in the short term this consideration can be outweighed by other influences such as interest rates, trade flows and political stability. You should also remember that as in other markets, buyers and sellers are as much concerned with the future as with the present and the past. If the market thinks that a currency is likely to fall in the future, it will anticipate that belief by selling now so that expectations can be self-fulfilling. This does not mean that the market is always right. Anticipations about future movements are based on past experience, so that the market may not recognise that a fundamental shift has taken place until this becomes completely clear and then it may overreact.

For example, between 1962 and 1992 Britain had a generally poor record in controlling inflation. By 1995 currency markets remained sceptical about future inflation rates in Britain, in spite of the declared intentions of the British government and its relative successes between 1992 and 1995. Over a similar period Japan's economic record had been one of spectacular success, so that the market continued to believe that its economic problems of the first half of the 1990s were likely to be temporary. It is quite feasible that the judgement of the currency markets was wrong in the mid-1990s for both countries. The currency traders risked losing a great deal of money if their beliefs were wrong and only future events will show whether or not they were correct.

Exchange Rate Structures

There are basically two types of exchange rate system – *fixed* and *floating* exchange rates. There may be variants on these, but the basic principles remain the same.

(a) Fixed Exchange Rates

It is very rare to have an exchange rate structure that is rigidly fixed. Some movement within a band either side of a central rate is normal. The more confident governments are that they can maintain the agreed rates, the narrower the band within which floating is permitted. A movement towards either the floor or the ceiling of the band requires action to correct the rate. The usual short-term action is to change interest rates to attract – or discourage – capital movements, but longer-term action through taxation or a fundamental shift in government spending or policy priorities is likely to be needed. If the government is unable or unwilling to take action to restore the agreed exchange rate, or if its action is unsuccessful, then the rate will have to be changed. If member countries cannot agree on a satisfactory change the whole structure becomes unstable.

The problem with any fixed exchange rate structure is reconciling the desired level of stability with sufficient flexibility to allow changes to take place as economic conditions change. National economies are dynamic. They are subject to constant change. A system designed to prevent short-term fluctuations can easily block desirable long-term developments, until the currency values get so out of touch with reality that a structural upheaval becomes inevitable.

Nevertheless there have been a number of important attempts to create exchange rate structures to provide the stability that business firms desire.

© ABE and RRC

The longest, most comprehensive and for many years the most successful attempt was the Bretton Woods system (see Study Unit 15). This linked the main currencies to the United States dollar throughout the 1950s and 1960s – a period of generally rising world living standards and of considerable prosperity for the Western world. The European Community's Exchange Rate Mechanism (ERM) sought to reproduce the Bretton Woods conditions. It had a roughly similar system of limited currency movements within defined bands, and operated during the 1980s and 1990s in the lead up to the establishment of the single European currency. Supporters of such systems usually claim that they:

- provide the stability and reduction in currency risks that traders need if they are to expand trade and production
- oblige governments to pursue financially responsible economic policies designed to control inflation and curb the tendencies of communities to live beyond the means provided by their production and trading systems.

Opponents of fixed rate structures point out that periods of apparent exchange rate stability tend to be punctuated with intense speculative crises and periods of serious and damaging instability. This happens when finance markets realise that a major currency (usually sterling!) has become overvalued and they suspect that the government does not have the power to prevent a devaluation. A series of crises led to the abandonment of the Bretton Woods system in the early 1970s and a similar crisis led to the withdrawal of sterling from the ERM in 1992.

Opponents also point out that the only measures that governments can take to uphold the exchange value of a currency in the short term are extremely damaging to their domestic economies and further undermine long-term confidence in the currency. A monetarist government will rely on high interest rates to keep capital in the country, but these high rates can have a devastating effect on consumer demand and business investment, as shown in Britain in the period 1989–1992. A Keynesian government would raise taxation and curb wages and other incomes, and this would have a similar deflationary effect to high interest rates. Clearly a government seeking to maintain an overvalued currency will damage its own domestic economy, create high unemployment and destroy business firms. Living standards fall in the interests of an artificial currency stability, which cannot be sustained for more than a short period.

Currency exchange rates represent the market price of a nation's currency. They are the international traders' valuation of the nation's production system. Stable exchange rates can only be achieved when economies are themselves stable, prosperous and competitive in world markets. A falling exchange rate is the symptom of an unhealthy economy. To prop it up artificially is like propping up a weak patient and pretending that the patient is fit and well. It is as dangerous to the economy as it is to a sick person, and eventually all such pretences have to be abandoned.

(b) Floating Exchange Rates

When the price of the currency in terms of every other currency is set by demand and supply in the market, the country is said to have a freely floating exchange rate. If the demand increases and the supply remains the same, the exchange rate rises (*appreciates*); should the supply increase faster than demand, the rate falls (*depreciates*). There are no exchange controls and the government does not intervene in the market. Figures 17.1 and 17.2 show how changes in demand and supply affect the exchange rate of a currency.

© ABE and RRC

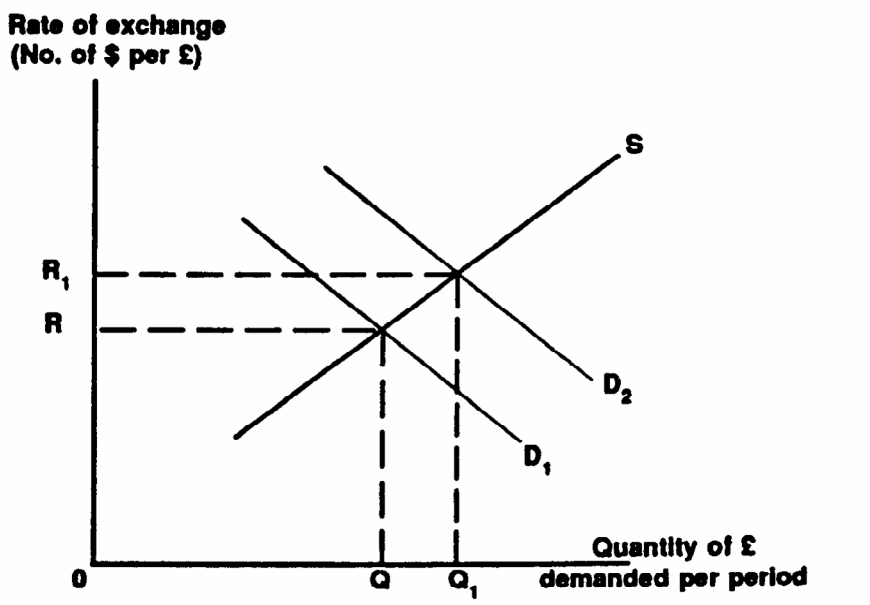


Figure 17.1: The effect of increased UK exports or more investment in Britain

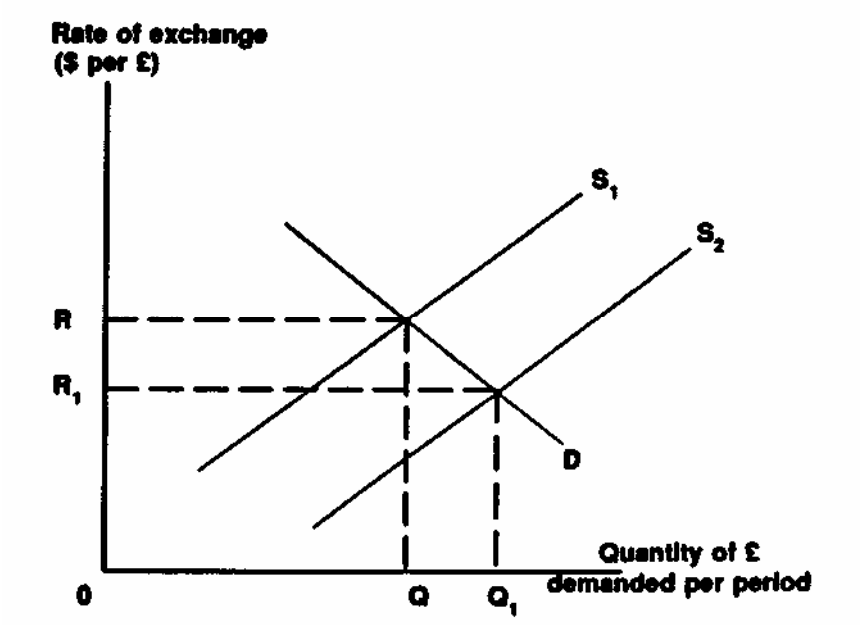


Figure 17.2: The effect of increased UK imports or more UK investment abroad

If Britain's exports increase there will be more demand from importers to exchange their currencies into sterling. The pound will also be in demand if people want to invest more in the UK, either in deposits and shares or in physical assets. More sterling will be supplied if importers in Britain are buying more from overseas and require more foreign currency. UK investment abroad increases the supply of pounds.

Just as in any other market, an increase in demand for pounds, with supply unchanged, will cause the price of sterling to rise or appreciate – more dollars have to be paid for each pound. Conversely an increase in supply, with demand remaining the same, would cause the currency to depreciate and each dollar would buy more pounds – i.e. the price of a pound has fallen.

© ABE and RRC

Governments have often attempted to manage floating exchange rates: this is called *"dirty floating"*. A government may intervene in the market to buy or sell its currency because it wants to hold down a rise in the rate, which would affect international competitiveness, or support a rate, to keep foreign investments.

There have been attempts by the major industrial countries to influence the exchange rate of the US dollar. Many commodities and raw materials, especially oil, are priced worldwide in dollars; a rise in the value of the dollar for speculative reasons unconnected to trade could cause inflation. When, in 1991, the dollar rose by a quarter against the Deutschmark, the G7 (the seven most industrialised nations) took concerted action to stem the rise by central bank intervention to sell dollars. In 1995 the dollar was falling against other currencies because of fears about the effect of the very large US government deficit and the political situation. This led to a flight into the Deutschmark, a rise in its rate and a depreciation of other currencies. The effect is to make the exports of appreciating countries less competitive and those of depreciating ones more so – this is destabilising and has nothing to do with the trading position of the countries. Central banks intervened to buy dollars in an attempt to prevent further falls in the rate.

Even when all the major central banks act together, they cannot have a significant effect on the foreign exchange market. The sheer size of the market's daily dealings makes the reserves of the industrialised countries look small. The banks can try to influence the feeling in the market so that dealers change their attitude to the future of the currency.

The advantages of floating exchange rates are:

- There is an inbuilt adjustment mechanism. If imports exceed exports, the currency will depreciate and exports become relatively cheaper in foreign countries, thus helping to increase exports. There is no need for government intervention.
- There is continuous adjustment of the rate, in contrast to the infrequent, large and disruptive revaluations in fixed systems.
- Domestic economic policy can be managed independently of external constraints imposed by the need to maintain the exchange rate.
- There is no possibility of imported inflation, as the exchange rate adjusts relative prices.
- There is no need for large official reserves (unless there is managed floating).
- Adjustments to the exchange rate are made by the market: they are not delayed by political considerations.

The disadvantages of floating exchange rates are:

- They create uncertainty and raise the costs of international activities because of the need to cover risk.
- There are no restraints on inflationary domestic economic policies.
- Changes in the rate may be due to speculation or flight from weakening currencies and have nothing to do with the trading position of the country. This may make exports relatively dearer and imports cheaper and cause a payments deficit.

The impact of a change in a floating exchange rate depends on the price elasticities of demand for exports and imports. If both are elastic, a fall in the rate will reduce imports, which become dearer in the home market, and increase exports, which become cheaper in foreign markets. The opposite happens if the rate appreciates. If

© ABE and RRC

the demand for exports abroad is inelastic, the effect of depreciation will be that the volume of exports does not increase but the lower price earns less foreign exchange. If imports are price inelastic, the rise in their price does not reduce demand significantly and more foreign exchange is bought to pay for them: this worsens the balance of payments. Higher import prices for materials, components and finished goods may cause inflation.

C. EXCHANGE RATE POLICY

Exchange rate policy refers both to a country's choice of exchange rate regime and its use of its exchange rate to achieve its macroeconomic policy objectives. In the late 1940s and most of the 1950s exchange rate policy would have been largely focused on the decision whether to adopt a rigidly fixed exchange rate regime or allow a country's currency to float freely.

A *freely floating exchange rate* enabled a government to use monetary and fiscal policy measures to achieve the internal objectives of macroeconomic policy, without the constraint of worrying about its external balance of trade position.

On the other hand, a *fixed exchange rate* regime was seen as beneficial to the promotion of international trade, because it removed exchange rate uncertainty from importing and exporting activities. A commitment to fixed exchange rates also reflected the desire to avoid using frequent exchange rate devaluations as a means of attempting to gain unfair advantage from international trade.

Frequent changes in exchange rates led to competitive devaluations and damaging trade wars in the 1930s. Reflecting on this experience, which led to a collapse of international trade and merely served to spread unemployment around the world rather than the benefits from trade, countries favoured fixed exchange rates with the formation of the International Monetary Fund in 1945.

More recently, the choice of exchange rate regime has been recognised to exert a big influence on the relative effectiveness of monetary and fiscal policy. In addition, the choice of a fixed exchange rate regime means that a country loses the ability to determine its own rate of inflation, and must accept that it will experience a rate of inflation determined by the rest of the world. In contrast, the choice of a freely floating exchange rate means that a country is in control of its own rate of inflation because its nominal exchange rate will adjust to isolate it from the world rate of inflation. (Go back to Study Unit 16 and revise your understanding of purchasing power parity if you do not understand how this process works). Thus, if a government wants to achieve a low rate of inflation as its main objective of macroeconomic policy, it is likely to favour a freely floating exchange rate regime.

The other aspect of exchange rate policy has to do with the objectives of achieving full employment and a high rate of economic growth based on exporting; this is referred to as *export led growth*, and involves the *terms of trade*. We introduced the concept of the terms of trade in Study Unit 16. To recap, the terms of trade measures the relative movement of import and export prices. It is calculated from:

$$\frac{\text{unit value index of exports}}{\text{unit value index of imports}} \times 100$$

The unit value index represents the average movement in price of a "unit" of imports or exports. The unit itself is a kind of average of all types of visible imports and exports. The terms of trade thus gives a general indication of how average import and export prices are moving. A high terms of trade is beneficial for a country, provided it goes hand in hand with a high demand for its exports. But a high terms of trade also results from overvaluation of a country's currency, and if this leads to falling exports and rising imports the country will suffer. A country can manipulate its exchange rate to alter its terms of trade.

© ABE and RRC

A country may adopt a fixed value for its currency that is deliberately undervalued, so that its export industries have a big competitive advantage in international markets. This policy will worsen its terms of trade and make imports expensive, but it can lead to export led growth and a very large surplus on its balance of trade. The low terms of trade means that the country suffers a lower standard of living than it could achieve if it increased its exchange rate, or allowed its currency to appreciate. This is because it is selling its exports "cheaply" in international markets relative to what it has to pay for its imports. But on the plus side, if its exchange rate is sufficiently undervalued as to give its firms a really big cost advantage in exporting, and it can resist the pressure from those countries experiencing huge trade deficits as the counterpart of its huge trade surplus to revalue its currency, then its industry, employment and growth will prosper. The best example in recent times of a country deliberately maintaining an undervalued fixed exchange rate to boost its economic growth is provided by the rise to dominance of China as one of the world's leading export nations. Such a policy does not come without its economic consequences. As explained next, maintaining a fixed exchange rate leaves a country open to importing inflation. Artificially depressing the terms of trade to gain an advantage in exporting adds further to domestic inflationary pressure by increasing the price of imports. This is the problem experienced by China towards the end of the first decade of the twenty-first century.

China is not the first or only country to seek to grow its domestic economy through export-led growth based on maintaining an undervalued currency. The best example is provided by Japan. Japanese economic policy towards its exchange rate under the IMF Bretton Woods system of fixed exchange rates was to keep its currency seriously undervalued, and resist all pressure, especially from its main export market in the USA, to revalue its currency. Japanese success as one of the world's leading exporters owes much to its exchange rate policy. Since Japan adopted a floating exchange rate in the 1970s, the Japanese government and the Bank of Japan have managed the exchange rate through intervention in the foreign exchange market, to limit its appreciation and maintain Japanese companies export competitiveness. The extent of the intervention is seen most clearly whenever the yen appreciates against the US dollar and looks like increasing to such an extent that the US dollar falls below 100 yen to the dollar. When this happens the yen soon loses value again and depreciates in value against the US dollar, much to the relief of Japanese based exporters.

D. MACROECONOMIC POLICY IN OPEN ECONOMY

In Study Units 13 and 14 we explained, using both the Keynesian 45 degree model and the aggregate demand and supply model of income determination, how governments could use monetary and fiscal policies to influence the level of demand in the economy and achieve the objectives of macroeconomic policy. In the analysis of income determination we allowed for exports as an injection of aggregate demand and imports as a withdrawal of aggregate demand from the economy, but neglected the economy's exchange rate regime. This was done to simplify the analysis and make the exposition easier to follow. However by ignoring the type of exchange rate operated by a country we have overstated the effectiveness of monetary and fiscal policies and the power of a government to control the economy.

Economics teaches us that there are some things that are beyond the control of governments. For example, when the demand for a good or service increases its price will rise, unless the increase in demand is matched by an equal increase in supply. The rise in price may be unpopular but it is unavoidable, because no government can abolish scarcity, and the laws of economics, by decree. The same applies to macroeconomic policy. It can be proved (but will be simply stated here to avoid a long and complex piece of analysis) that a *government cannot control all three of the following macroeconomic variables at the same time: the rate of interest, the exchange rate and the rate of inflation.*

© ABE and RRC

Governments face a dilemma or policy conflict when it comes to choosing between these three variables. They can only choose to determine the value of one of the three as a policy objective or target. Once they have fixed the value of one of the three, the values of the other two variables will be determined by market forces. Thus if a government decides to fix the value of its currency against that of another country by adopting a fixed exchange rate regime, the government will have to accept that it cannot also determine the level of interest rates in the economy and control the rate of inflation. Rather, the government will have to vary the rate of interest to defend its fixed value of its exchange rate, and how it changes the level of its rate of interest will be dictated by rate change overseas. Likewise, the rate of inflation in the country will be determined partly by the level of interest rates and the rate of inflation in the global economy. If a government decides that its most important macroeconomic policy objective is to control the rate of inflation, then it must sacrifice its ability to simultaneously determine its exchange rate and level of interest rates.

This particular dilemma explains why most of the world's advanced economies have abandoned fixed exchange rates in favour of floating exchange rates, and given their central banks independence to use interest rates to achieve a fixed target for the rate of inflation. Given the choice between a fixed exchange rate and achieving a target rate of inflation, many governments have decided that a floating exchange rate is a small price to pay for achieving control over the rate of inflation. Conversely, those countries that have opted to operate a fixed exchange rate regime for trade advantage reasons, especially China, have discovered the hard way that eventually this policy choice leads to the problem of increasing domestic inflation.

An open economy enables a country to enjoy the gains from international trade, but it also constrains the choice of macroeconomic policy objectives. There is a further consequence: the choice of exchange rate regime also affects the effectiveness of monetary and fiscal policies in controlling demand in the economy. Governments need to recognise that:

- Fiscal policy is most effective and monetary policy least effective if a country operates a fixed exchange rate regime.
- Monetary policy is most effective and fiscal policy least effective if a country operates a freely floating exchange rate.

The explanation for this involves the rate of interest. Remember that as the level of national income increases, so does the demand for money. If the supply of money remains constant, this will cause the rate of interest to increase. Remember also that increased borrowing by a government, to finance its budget deficit, will drive up the level of the rate of interest. If economies are open to international trade and financial flows, then differences in interest rates between countries will cause investing institutions to move funds between countries in search of the highest return. The flow of funds into and out of a country will result in pressure on its exchange rate to change. The implication of these relationships depends upon a country's exchange rate regime.

Consider a country operating a fixed exchange rate regime. The country's central bank will have to use the rate of interest and intervention in the foreign exchange market to maintain the exchange rate at the fixed level chosen by the government. If the government undertakes an expansionary fiscal policy, the resultant upward pressure on the rate of interest will attract an inflow of money from the rest of the world. If this is unchecked, it will cause the exchange rate to appreciate above its fixed rate value. This will force the central bank to intervene in the foreign exchange market, by buying foreign currency at the fixed rate and increasing the supply of the domestic currency. The increased supply of the domestic currency will put downward pressure on the rate of interest. The net result is that the expansionary fiscal policy is unchecked by any induced off-setting rise in interest rates. Fiscal policy is thus highly effective in this case. In contrast, monetary policy is largely ineffective under a regime of fixed interest rates. For example, an expansionary monetary policy will lower the domestic rate of interest and cause an outflow of funds from the

© ABE and RRC

economy. The outflow of the domestic currency increases its supply relative to demand on the foreign exchange market, and causes downward pressure on the exchange rate. To maintain the fixed value for the exchange rate, the central bank has to intervene in the foreign exchange market by selling foreign currencies from the country's reserves, and in return take domestic currency out of the market. The consequence of this buy back of domestic currency by the central bank is to push the domestic rate of interest back up to its value before the expansionary monetary policy was undertaken. The net result of the attempted expansionary monetary policy is that the domestic money supply and the rate of interest return to their initial values, but the country has a small stock of foreign currency reserves.

If a country operates with a freely floating exchange rate regime the previous conclusions regarding the effectiveness of fiscal and monetary policy are reversed completely. The value of the exchange rate is now determined by the forces of demand and supply in the foreign exchange market, without any intervention by the central bank. An expansionary monetary policy reduces the rate of interest and causes funds to flow overseas in search of a higher return. Without any intervention by the central bank, the increased supply of domestic money on the foreign exchange market will cause the currency to depreciate, i.e. the value of the exchange rate will be reduced. This depreciation of the exchange rate has two consequences which enhance the effectiveness of monetary policy in boosting demand. The depreciation of the currency will make exports more competitive, and thus boost the demand for the country's exports. The depreciation in the exchange rate also makes imports more expensive, and will cause domestic demand to switch from imports towards domestic suppliers. Both of these effects, the strength of which depends upon elasticity of demand and supply, increase injections and reduce withdrawals from the circular flow of income. This reinforces the initial boost to demand from the reduction in interest rates. Monetary policy is highly effective in this case. The same process works in reverse to strengthen the demand reducing effect of a contractionary monetary policy.

With a freely floating exchange rate fiscal policy is largely ineffective, because of the way in which it induces off-setting changes in the exchange rate. For example, an expansionary fiscal policy which initially boosts demand and causes the rate of interest to rise. The rise in the domestic interest rate relative to the level overseas will cause foreign demand for its currency to rise on the foreign exchange market and its value to appreciate. As the currency appreciates the country's export competitiveness will decline, and it will experience a decline in its exports. At the same time, the appreciation of the currency will make imports and overseas travel more attractive. Thus as the government's fiscal expansion increases injections into the circular flow of income, either in the form of more G, or C and I, the induced affect on the rate of interest and the exchange rate produces an off-setting decline in X and increase in M. Fiscal policy is thus rendered ineffective due to interest rate and exchange rate "crowding out".

This explanation is simplified, and in practice monetary and fiscal policy are never completely ineffective whichever exchange rate regime a country operates. This is because freely floating exchange rates are rarely left completely free by central banks, and funds are not completely free of all restrictions to move between all countries. However, the basic point remains valid. It helps to explain why, following the adoption of floating exchange rates by many governments from the 1970s onwards, much more importance is given to monetary policy to control the level of demand and hence the rate of inflation in an economy. Fiscal policy is still used to influence aggregate demand, but much less so than in the 1950s and 1960s, when most countries adopted a fixed exchange rate regime. Today fiscal policy is used more to achieve supply-side objectives rather than regulate aggregate demand in the economy.

© ABE and RRC

Review Points

You should go back to the start of this unit and check that you have achieved the learning objectives. If you do not think that you understand the aim and each of the objectives completely, you should spend more time rereading the relevant sections.

You can test your understanding of what you have learnt by attempting to answer the following questions. Check all of your answers with the unit text.

1. Explain the difference between devaluation/revaluation and depreciation/appreciation of currencies on the foreign exchange market.
2. What is purchasing power parity?
3. If a country has a higher rate of inflation than other countries then its nominal exchange rate will eventually depreciate to maintain purchasing power parity.

   True or false?
4. What is meant by the terms of trade?
5. Explain the meaning of "export led growth".
6. What are the advantages of a country choosing a freely floating rather than a fixed exchange rate?
7. Monetary policy is more effective than fiscal policy if a country chooses to operate a fixed floating exchange rate regime.

   True or false?

© ABE and RRC